DEPARTMENT OF CLINICAL NEUROSCIENCE
Karolinska Institutet, Stockholm, Sweden

MECHANISMS OF FLAVOR PERCEPTION - HOW ODOR AND TASTE INTERACT WHEN WE EAT

ROBIN FONDBERG

Stockholm 2021

Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2021

ISBN 978-91-8016-271-5

Mechanisms of flavor perception - How odor and taste interact when we eat
THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

ROBIN FONDBERG

The thesis was defended in public at Nobels Väg 9, Stockholm, 2021-06-23.

Principal Supervisor:
Associate Professor Janina Seubert
Karolinska Institutet
Department of Clinical Neuroscience
Division of Psychology

Co-supervisors:
Associate Professor Johan N. Lundström
Karolinska Institutet
Department of Clinical Neuroscience

Professor Mats J. Olsson
Karolinska Institutet
Department of Clinical Neuroscience

Opponent:
Dr. Maria Geraldine Veldhuizen
Mersin University
Department of Anatomy
Faculty of Medicine

Examination Board:
Associate Professor Linus Andersson
Umeå University
Department of Psychology

Professor Mats Nilsson
Stockholm University
Department of Psychology

Professor Paulina Nowicka
Uppsala University
Department of Food studies, Nutrition and Dietetics

Till Bob och Judit.

ABSTRACT

Every time we eat, our brains are bombarded with sensory information from the olfactory and gustatory modalities. Through a binding process that is not fully understood, the odor and taste are then merged into flavor, a unitary sensation that appears to arise from inside the mouth. Flavors are used to guide our behavioral responses to potential sources of nutrition: flavorsome foods can easily be consumed in excess, while distasteful foods are generally rejected.

Recent studies have shown that familiar flavors with congruent odor and taste components are perceived and evaluated differently than less familiar flavors with incongruent components. In this context, congruency denotes the extent to which an odor and a taste are associated with the same food object. A citrus odor mixed with a sweet/sour taste is an example of a congruent combination, because both sensations are associated with lemons. By contrast, chicken odor and sweet/sour taste would by most people be perceived as an incongruent combination.

That congruency affects flavor perception suggests that associative learning within the olfactory-gustatory network regulates the binding process that gives rise to flavor. Investigating how odors and tastes that are frequently encountered together become mentally linked to one another has the potential to advance our understanding of how food preferences are formed and evolve over time. In this dissertation, I will present three lab experiments that in different ways have explored the role of associative learning in flavor perception.

Study I (*n*=30) and **Study II (*n*=23)** investigated perceptual and hedonic effects of learning that has already taken place. In these studies, participants rated several flavors with varying degrees of congruency. Congruency was manipulated in a linear fashion: some flavors were highly congruent (e.g. citrus odor+sweet/sour taste), some were highly incongruent (e.g. chicken odor + sweet/sour taste), and others were moderately congruent (e.g. chicken/lemon odor mixture + sweet/sour taste). **Study I** first showed that congruency has a positive, linear-like effect on flavor pleasantness. The more congruent the particular odor-taste mixture, the more pleasant the flavor sensation. This study also provided weak evidence that congruency increases the probability that a flavor's odor will be referred (or mislocalized) to the mouth, a perceptual illusion that has been suggested to reflect that the unisensory components have been bound together as a unified and meaningful whole.

Previous research has shown that hunger makes food more appetizing. By adhering to a pre-registered analysis plan, the primary aim of **Study II** was to test whether the amplifying effect of congruency on pleasantness interacts with the hunger state of the perceiver. To promote consumption of familiar (and safe) foods, we expected the amplifying effect of hunger to be stronger on congruent than on incongruent flavors. Participants attended two experimental sessions spaced approximately one week apart. One session was completed during hunger and the other session was completed during satiety. This study first replicated the positive effect of congruency on pleasantness from **Study I**. However, contrary to

expectations, the congruency by hunger state interaction was not significant. Although **Study II** provided no evidence that the effect of hunger on flavor pleasantness is stronger for congruent than for incongruent flavors, this fining should be considered preliminary due to the small sample size.

Taken together, **Study I** and **Study II** provide strong evidence that congruent flavors are more appetizing than incongruent flavors (at least for some odor-taste combinations). This suggests that frequently encountered foods that have been determined through experience to be safe are preferred to novel foods that are associated with greater risks of negative metabolic consequences. An exploratory analysis of the combined datasets from **Study I** and **Study II** showed that the amplifying effect of congruency was linear. This suggests that while congruent flavors indeed are most pleasant, minor differences between what is perceived (a specific odor-taste mixture) and what is expected (a perfect prototype of the encountered food item) will likely be tolerable. Such graded response pattern may underlie our ability to accommodate fluctuations in the chemical composition of different foods.

Study III (*n*=60) was designed to create new odor-taste associations. In this pre-registered study, two relatively unfamiliar odors were repeatedly presented during a five-day exposure phase. One odor was presented with a sweet taste and the other odor was presented alone. Four outcomes that are thought to be affected by associative learning were rated before and after the exposure phase: odor sweetness, odor pleasantness, odor intensity enhancement by taste, and odor referral to the mouth. Contrary to expectations, repeatedly presenting odor and taste together had no effect on any of the outcomes. Moreover, exploratory equivalence tests suggested that the effects were either absent, or substantially smaller than in previous studies. High-powered, transparently conducted, direct replications of studies with significant results are needed to confirm that associative learning effects can reliably be observed in experimental settings. If this turns out to be the case, follow-up studies should focus on identifying contextual factors that modulate these effects.

In this dissertation, **Study I**, **II**, and **III** will be discussed in light of the so-called replication crisis (or credibility revolution) in experimental psychology. Methodological advancements that can be adopted to increase the trustworthiness of flavor research are highlighted, together with some recommendations on how the field should proceed to determine what associative learning actually contributes to the eating experience.

LIST OF SCIENTIFIC PAPERS

I. **Fondberg, R.,** Lundström, J. N., Blöchl, M., Olsson, M. J., & Seubert, J. (2018). Multisensory flavor perception: The relationship between congruency, pleasantness, and odor referral to the mouth. *Appetite*, *125*, 244-252. doi: 10.1016/j.appet.2018.02.012

II. **Fondberg, R.,** Lundström, J. N., & Seubert, J. (2021). The relationship between hunger state, odor-taste congruency, and pleasantness evaluation of flavor. [Unpublished manuscript], Department of Clinical Neuroscience, Karolinska Institutet

Open materials: https://osf.io/n7wyc/, doi: 10.17605/OSF.IO/N7WYC
Preregistration: https://osf.io/p7jus, doi: 10.17605/OSF.IO/P7JUS

III. **Fondberg, R.,** Lundström, J. N., & Seubert, J. (2021). Odor-taste interactions in food perception: Exposure protocol shows no effects of associative learning. *Chemical Senses*, *46*.
doi: 10.1093/chemse/bjab003

Open materials: https://osf.io/dtv8s/, doi: 10.17605/OSF.IO/DTV8S
Preregistration: https://osf.io/p6ab3/, https://aspredicted.org/j6s5n.pdf

SCIENTIFIC PAPERS NOT INCLUDED IN THE THESIS

I. Armeni, K., Brinkman, L., Carlsson, R., Eerland, A., Fijten, R., **Fondberg, R.,** ... & Zurita-Milla, R. (in press). Towards wide-scale adoption of open science practices: The role of open science communities. *Science and Public Policy*. Preprint doi: 10.31222/osf.io/7gct9

II. Lalouni, M., Fust, J., Bjureberg, J., Kastrati, G., **Fondberg, R**., Fransson, P., Jayaram-Lindstrom, N., Kosek, E., Hellner, C., Jensen, K. J. (2021). Central inhibition of pain is augmented in women with self-injurious behavior. [Unpublished manuscript]. Preprint doi: 10.1101/2021.05.12.21257091

CONTENTS

KEY CONCEPTS

Olfaction	The sense of smell
Odor	A sensation resulting from stimulation of the olfactory modality
Retronasal odor	An odor that is generated by odorous molecules that have reached the nasal cavity via the mouth and throat
Orthonasal odor	An odor that is generated by odorous molecules that have reached the nasal cavity via the nostrils
Odorant	A substance that has a distinctive smell
Gustation	The sense of taste
Taste	A sensation resulting from stimulation of the gustatory modality
Tastants	A substance that has a distinctive taste
Flavor	The combined sensation of odor and taste
Congruency	The extent to which an odor and a taste are associated with the same food object

1 INTRODUCTION

Eating is essential for survival, but also one of the most enjoyable activities in life. During each meal, synchronized activation of the olfactory and gustatory modalities results in complex sensations that are experienced inside the mouth (Small 2012; Spence 2015b; Stevenson 2014). These sensations are commonly referred to as flavors. Becaus the gustatory and olfactory modalities are tuned to different types of chemical compounds, flavors carry much more information than could be obtained by any of the unisensory modalities alone. This feature makes them ideal for promoting informed decision-making during both food selection and ingestion.

In terms of biological relevance, flavor perception forms a central part of a broader system that functions to maximize nutrient intake while minimizing the risk of food poisoning (Stevenson 2009). The first sip of a perfectly flavored hot chocolate will leave you craving for more, an effect that strongly promotes consumption of fast carbohydrates. By contrast, the same drink will probably be rejected if it tastes sour, a defense mechanism that protects you from getting sick from ingesting spoiled milk. However, the same sour taste can be pleasant when consumed in a different sensory context. Lemonade, for example, is by most evaluated as pleasant despite its acidity. This demonstrates that our hedonic responses to foods are not only modulated by the unisensory odor and taste components, but also by the actual composition of the flavor mixture.

There is little doubt that exposure plays an important role in the development of flavor preferences (Schifferstein and Verlegh 1996; Small et al. 2004). Hedonically neutral odors-taste combinations that are repeatedly consumed without aversive consequences are thought to eventually become pleasant (e.g. Hausner, Olsen, and Møller 2012). This positive hedonic shift will stimulate consumption of familiar foods that have been determined through experience to be safe and nutritious. Interestingly, repeated exposure to flavors also seems to affect how the unisensory components are perceived individually. For example, the perceived intensity of a strawberry odor will increase when presented with a sweet taste, however, intensity will not be affected by the presence of a salty taste (Green et al. 2012). Several lines of research have presented converging evidence that odor-taste congruency, i.e. the extent to which the unisensory components are associated with the same object, affects both hedonic and perceptual characteristics of flavor. However, the exact effects of congruency, as well as the learning process through which congruency develops due to repeated exposure to olfactory-gustatory mixtures, remain poorly understood.

Because olfaction and gustation are the building blocks of flavor, this introduction will first focus on how these unisensory modalities function in isolation (*1.1 - Flavor components*). After that, the multisensory binding process that combines odor and taste into flavor will be described (*1.2 - Flavor binding*). The next section summarizes the literature that has explored the potential role of associative learning in this binding process (*1.3 - Associative learning*).

During the last decade, the so-called replication crisis has exposed severe problems in many social and biomedical fields. As a consequence, researchers have started to question the credibility of many findings in experimental psychology. This has obvious implications for chemosensory research where small samples are the norm and replication studies specifically designed to confirm published claims are scarce. The last section of this introduction describes this development (*1.4 – Replication crisis*).

1.1 FLAVOR COMPONENTS

1.1.1 Gustation

While the layman tends to attribute most of what is experienced during a meal to the gustatory modality (Rozin 1982), pure taste contributes surprisingly little to flavor (Spence 2015a). Only five taste qualities are indisputably accepted as basic tastes (Chandrashekar et al. 2006; for evidence that humans may detect more than five tastes see Iwata, Yoshida, and Ninomiya 2014) each of which has a specific role for detecting compounds with high biological relevance. Three of them signal nutritious compounds (sweet for carbohydrates, salty for salts, and savory for proteins), and two signal potentially harmful compounds (sour for acids and bitter for poisons). There is definitely variability both between individuals and specific tastants but, in general, the hedonic tone of a taste depends on the evolutionary benefit of the compound it is associated with. Several lines of research suggest that while the gustatory modality only contributes a handful of qualities to flavor, it is absolutely central for assigning motivational value to food objects in our environment (Beauchamp and Mennella 2011).

A few days after birth, human infants show robust taste preferences according to the nutritious/harmful divide. Sweet and savory tastes are liked, while bitter taste is disliked (Ventura and Worobey 2013). These hedonic responses indicate that the gustatory system may be hardwired to meet the evolutionary need to consume nutritious compounds while avoiding toxins. During food consumption, the hedonic tone of the taste is thought to change the hedonic tone of the odors it is presented with (Mura Paroche et al. 2017). Following this associative learning process, the odors themselves will start to evoke similar hedonic responses as the taste. Those learned responses let us infer hedonic properties of foods based on their smells alone, an ability that will protect our internal environment by enabling us to anticipate the nutritive value of different foods without having to taste them.

Taste signaling is initiated by small populations of specialized taste cells grouped in clusters, taste buds, which are dispersed throughout the tongue. Isolated taste buds are found in the soft palate, the pharynx, and the esophagus. When tastants bind to receptors embedded in the membrane of a taste cell, neurotransmitter molecules are released across the synapse which initiates an action potential along one out of three cranial nerves (facial, glossopharyngeal, or vagus). This taste signal is relayed to the solitary nucleus in the medulla oblongata (Roper and Chaudhari 2017; Witt, Reutter, and Miller, Jr. 2003), and from there, to the ventral posterolateral nucleus of the thalamus. After synapsing, the signal finally reaches the insula with its overlying operculum, which together constitute the primary gustatory cortex (Smith and Boughter Jr. 2009). Recent evidence suggests that neuronal response patterns in the insula can distinguish between sweet, salty, sour, and bitter (Chikazoe, Lee, Kriegeskorte, and Anderson 2019; Crouzet, Busch, and Ohla 2015; Porcu et al. 2020). This means that quality may be the first perceptual feature of taste that is represented during gustatory processing. While conscious taste percepts are thought to rely on activation in these primary regions (Small et al. 1999; Veldhuizen et al. 2011), the hedonic dimension of taste is thought to be encoded in the orbitofrontal cortex, which is a secondary gustatory area (McCabe and Rolls 2007).

How the nervous system discriminates between different taste qualities is not yet fully understood. Three models have been proposed to explain the mechanisms behind gustatory quality coding (Roper and Chaudhari 2017). The labelled line coding model is the one that has the most support. This model states that each taste quality has its own specialized signaling pathway that begins at the receptor level and ends in the gustatory cortex. The function of each individual taste cell and afferent fiber in the system is to signal one particular taste quality (Chen et al. 2011). In support of this notion, many taste cells are indeed most sensitive to a single taste quality, and the axons of these cells remain separate all the way to the brain (Kinnamon and Finger 2019).

This model does, however, not tell the whole story. While some taste cells do respond most strongly to one specific taste quality, others have a much broader receptive field. For example, taste cells that detect sour can also respond to other taste qualities through communication with other taste cells (Roper and Chaudhari 2017). In addition, higher up in the gustatory processing stream, some afferent ganglion and hindbrain neurons are sensitive to several taste qualities. These observations support the combinatorial processing model, which states that information about taste quality is encoded in the combination of fibers that are activated at any given time point (Wu et al. 2015). One last possible mechanism for transduction is presented by the temporal patterning model, which states that it is the different temporal patterns of neuronal activity that code for taste quality (Roper and Chaudhari 2017). All three models continue to be raised as potential explanations of how the information is translated from taste receptors to percept, importantly, these models are not mutually exclusive and may play different roles in different parts of the gustatory processing stream.

1.1.2 Olfaction

In contrast to tastes, humans can perceive an impressively large number of odors. Although the exact number is hard to estimate (Gerkin and Castro 2015), most experts believe that we can discriminate between thousands of olfactory qualities. This ability explains why the odor component of flavor carries almost all the information required for food identification during meals. In fact, olfactory content normally dominates our eating experience (Spence 2015a), which most people seem surprisingly unaware of (Hollingworth and Poffenberger Jr. 1917; Rozin 1982).

In the context of food, it is relevant to highlight the dual nature of the olfactory modality. As the olfactory receptors are positioned in the nasal cavity (Schuenke and Schulte 2011), airborne odorants can reach their destination by two separate routes. Retronasal olfaction detects odorants emanating from food inside the mouth, which have reached the receptors via the throat. By contrast, orthonasal olfaction occurs when sniffing things in our external environment. There is some evidence that odors are perceived (Hummel et al. 2006) and processed (Hummel and Heilmann 2008) differently depending on the route of stimulation. While orthonasal olfaction surely contributes to the preparatory phase of eating (most foods are sniffed before they are tasted), only retronasal olfaction contributes directly to flavor.

Olfaction is initiated when odorants bind to receptors situated inside the olfactory epithelium on the superior part of the nasal cavity. These receptors belong to neurons whose axons constitute the olfactory nerve. Unlike gustation, which is supported by three separate nerves, olfaction thus relies on a single pathway to carry all the sensory information. The information flow may appear simple, but making sense of olfactory input is a huge challenge for our brains. Humans express about 400 different receptor proteins (Mainland et al. 2014). In addition, most odorants activate several olfactory receptors, and most olfactory receptors are activated by many different odorants (Bear, Connors, and Paradiso 2007). Each individual receptor type only provides a small fraction of the information needed for odor identification. Instead, it is the collective activation pattern from several receptor types that is interpreted along the central olfactory pathway and that eventually gives rise to unitary odor percepts.

The glomerular layer of the olfactory bulb, a small structure on the inferior side of the cerebral hemispheres just superior to the nasal cavity (Schuenke and Schulte 2011), is the first part of the brain that receives the olfactory input. Notably, neurons that express the same receptors will converge onto the same glomerulus (Firestein 2001), which gives rise to a spatial organization in the bulb that corresponds to the chemical properties of the odorants. In line with this observation, evidence from the animal literature suggests that the encoding pattern in this part of the brain is chemotopic (Johnson and Leon 2007), which means that the spatial arrangement of neurons mirrors the chemical attributes of the odorants. The signal is transmitted from the bulb to the piriform cortex, a key olfactory region that mediates identification, categorization, and discrimination (Gottfried 2010). The piriform cortex is connected to the insula, the orbitofrontal cortex, and other regions associated with emotion and memory (Freiherr 2017).

1.2 FLAVOR BINDING

Both behavioral and neuroimaging studies (Prescott 1999, Small 2012) have during recent years provided compelling evidence that flavor is something more than just the mere addition of its unisensory components. This literature has been inspired by the observation that during food consumption, humans do not perceive odor and taste as separate sensory events. Instead, a flavor is experienced as a unitary sensation that possesses both olfactory and gustatory qualities.

Results from these studies have shown that certain tastes can modify perceptual features of simultaneously presented odors and vice versa (Frank and Byram 1988; Green et al. 2012; Lim and Johnson 2011; Schifferstein and Verlegh 1996). These findings indicate that olfaction and gustation interact when we eat. Insights into the unique features of the multisensory percepts that result from this interaction, and the discovery of neural circuits where flavor-induced activation exceeds the summed activation evoked by the unisensory components (Seubert et al. 2015; Small et al. 2004), support the existence of a functionally distinct flavor modality (Gibson 1966). This modality has the unique ability to integrate information detected by sensory receptors that are anatomically separated in the peripheral nervous system.

The function of flavor encoding is to create hedonic memory representations that can be used to guide subsequent decisions about what to eat and what to reject (Spence and Piqueras-Fiszman 2014). During each meal, encountered sensations are thought to be compared to memory representations of similar foods. Evaluation of the flavor in relation to encoded memories will provide valuable information about the nutritional value of what is being consumed. Using flavors to represent foods has the advantage of combining gustatory information, which reliably signals macronutrients, with olfactory information, which possesses a broad sensory repertoire. This binding process thus results in the formation of percepts that uniquely can identify a broad range of foods, and whose biological relevance could not have been estimated by either of the two unisensory modalities alone.

1.3 ASSOCIATIVE LEARNING

1.3.1 Indirect support

Associative learning is thought to play a key role in flavor perception. The idea is that when unfamiliar food enters the mouth, a cortical process is initiated that creates an associative link between odor and taste components. If the exposure continues, these associations will eventually result in a robust flavor memory that modulates subsequent interactions between the unisensory components. Many observed characteristics of flavor are thought to be the result of this modulation (Prescott 2012; Small 2012). Moreover, once a flavor memory has been firmly encoded, presenting its odor components in isolation can sometimes be enough to activate the complete flavor memory (Prescott 1999). This explains why many odors commonly presented in desserts smell sweet (e.g. vanilla), and also why the perceived sweetness of an odor is a good predictor of how much it will enhance the rated sweetness of sucrose when presented together as flavor (Stevenson, Prescott, and Boakes 1999).

Because synchronized exposure strengthens the associative bond between a flavor's odor and taste, it will also result in increased perceived congruency. For this reason, congruency can be used as an indicator of how often the unisensory components have been perceived together in the past. In recent years, the importance of congruency in flavor perception has started to be explored. In a typical experiment, perceptual or hedonic ratings of a congruent (e.g. strawberry odor and sweet taste) and an incongruent (e.g. chicken odor and sweet taste) flavor mixture are compared in a relatively small number of healthy participants. This literature has revealed at least three features of flavor that can be modified by congruency.

The perhaps most interesting finding is that congruent flavors are more pleasant than what would be expected based on the hedonic value of their unisensory components (Schifferstein and Verlegh 1996; Small et al. 2004), which suggests that repeated exposure affects food preferences. Moreover, congruency also affects the perceived location of the odor component. It has long been known that when we eat, olfactory content is referred to the mouth (Hollingworth and Poffenberger Jr. 1917). To illustrate this concept, imagine eating an apple. The apple odor will be perceived in the mouth together with the sweet taste, rather than in the nose where the olfactory receptors are situated. Summarized under the term odor referral, this phenomenon has been suggested to at least partially explain why humans so often attribute olfactory content to the sense of taste (a sip of coffee is for example said to taste, not smell, like coffee; Spence 2016). Evidence that odor referral may be enhanced by congruency comes from studies by Lim and colleagues (Lim et al. 2014; Lim and Johnson 2011, 2012). In their experiments, odor referral was enhanced by the presence of congruent tastes, but not by incongruent tastes. Last, congruency may also affect the perceived intensity of the unisensory odor and taste components. Whereas flavors are perceived as unitary sensory events, their odor and taste qualities remain recognizable. An apple does for example have a sweet taste and a fruity aromatic profile, which demonstrates that both the gustatory (sweet) and olfactory (fruity) content are accessible to the perceiver. Adding a congruent tastant to an odor solution, e.g. sucrose to an unsweetened lemon drink, will amplify the intensity of the odor. By contrast, adding an incongruent tastant to the solution will not affect odor intensity (Fujimaru and Lim 2013; Green et al. 2012; Lim et al. 2014). This phenomenon is in line with findings from the literature on multisensory integration, which have revealed that the senses work together to enhance the salience of objects in our environment that are biologically relevant (Stein and Stanford 2008). It therefore makes sense that intensity enhancement only occurs when the odor and taste actually represent the same food.

Taken together, there is evidence that congruency affects at least three aspects of the flavor experience: flavor pleasantness, odor referral to the mouth, and perceived intensity. Given that congruent odor-taste combinations are combinations that belong to the same food object and therefore can be assumed to have been repeatedly perceived together, these results also provide indirect support that associate learning plays a role in flavor perception.

1.3.2 Direct support

Studies that have assessed effects of congruency to investigate the importance of learning in flavor perception have adopted a cross-sectional approach, comparing responses to odor/taste mixtures whose levels of congruency have already been established. More direct support for associative learning comes from studies that have presented initially unfamiliar odor-taste combinations repeatedly to the same participants. To isolate effects of associative learning that arise through synchronized activation of the olfactory and gustatory modalities from effects of exposure to unisensory stimuli, these studies have typically used two odors and one taste. One of the odors is presented together with the taste, and the other odor is presented alone. Perceptual and hedonic ratings are collected before and after the exposure. This approach have allowed researchers to compare changes between sweet-paired odors and odors that have been presented alone. Results from these studies suggest that odor sweetness (Prescott, Johnstone, and Francis 2004; Prescott and Murphy 2009; Stevenson, Boakes, and Prescott 1998; Stevenson, Prescott, and Boakes 1995; Stevenson and Case 2003), and possibly odor pleasantness (Baeyens et al. 1995; Stevenson and Case 2003; Yeomans, Prescott, and Gould 2009, but see also van den Bosch et al. 2015; Stevenson et al. 1998) and intensity (Stevenson et al. 1998, 1995, but see also van den Bosch et al. 2015; Labbe and Martin 2009; Yeomans et al. 2009), do increase more after exposure with sweet taste than after exposure without taste.

1.4 REPLICATION CRISIS

As tempting as it may be to interpret these studies as definitive proof that associative learning plays a key role in flavor perception, I think that every experimental psychologist working in a field where small, non-transparent studies of varying quality are the norm should interpret this type of evidence with some caution.

Over the last decade, many researchers have started to realize that some commonly used methodological practices may have created a literature that suffers from low replicability (obtaining the same result with new data; Open Science Collaboration 2015), reproducibility (obtaining the same result with the same data and the same analysis; Obels et al. 2020), and robustness (obtaining the same result with the same data but with different analyses; Kepes and McDaniel 2015). One event that contributed to the onset of the so-called replication crisis was the publication of an article by Bem in 2011 entitled "Feeling the future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect", which contained nine independent experiments that seemed to reveal that the future can causally affect the present. At first sight, the experiments appeared methodologically sound. The combined sample size was impressive (n>1000), data was analyzed using standard methods, and the article was published in a well-respected and peer-reviewed journal (Journal of Personality and Social Psychology, JPSP).

Despite the seemingly strong evidence for extrasensory perception, follow-up studies that closely mimicked the original failed to replicate the findings (i.e. Ritchie, Wiseman, and French 2012). This pattern of results would turn out to be surprisingly common. To date, several large-scale replications have been published that have repeated behavioral experiments published in peer-reviewed journals. When using statistical significance ($p<.05$) for evaluating replication success, approximately 50% have been successful (15/28 [Klein et al. 2018], 10/13 [Klein et al. 2014], 3/10 [Ebersole et al. 2016], 35/97 [Open Science Collaboration 2015], 11/18 [Camerer et al. 2016], 13/21 [Camerer et al. 2018]). Moreover, the replication effect sizes have been substantially smaller than the original effect sizes. Several lines of research suggest that different types of biases embedded in the academic infrastructure may contribute to the low replication rate. Two sources of bias have been particularly highlighted in the literature on replicability: publication bias and questionable research practices.

1.4.1 Publication bias

Despite sample sizes that on average are small, an overwhelming majority of published results are statistically significant in psychology research. In a paper from 1959, Sterling reviewed 362 articles published in four randomly selected journals and found that 97% claimed to have rejected the null hypothesis. A follow-up study based on results from articles published ~30 years later came to almost exactly the same conclusion (96% success rate; Sterling, Rosenbaum, and Weinkam 1995). More recent studies suggest that the situation today is not considerably different. Fanelli (2010) found that psychology and psychiatry had the highest percentage (92%) of positive results out of all assessed research disciplines, and the success rates of the original studies in the Reproducibility Project: Psychology (Open Science Collaboration 2015) and Scheel, Schijen and Lakens (2020) were 97% and 96%, respectively.

If this literature was free from bias and the reported proportion of positive results is a fair representation of the research that has been conducted, then both the proportion of all tested hypotheses that are true, and the average statistical power of the experiments that have tested those hypotheses, would need to exceed 90% (success rate = probability of the hypothesis being true * probability of detecting a true hypothesis). Such high numbers seem highly unlikely. Sterling (1959) writes:

> *There is some evidence that in fields where statistical tests of significance are commonly used, research which yields non-significant results is not published. Such research being unknown to other investigators may be repeated independently until eventually by chance a significant result occurs - an "error of the first kind" - and is published. Significant results published in these fields are seldom verified by independent replication. The possibility thus arises that*

the literature of such a field consists in substantial part of false conclusions resulting from errors of the first kind in statistical tests of significance.

This paragraph describes publication bias, the type of bias that occurs when positive findings have a higher probability of getting published than negative findings. Publication bias does not only result from decisions made by authors. It is common knowledge that "positive" results are easier to publish than "negative" results. For example, the study (Ritchie et al. 2012) that failed to replicate Bem's (2011) study on extrasensory perception was desk rejected by JPSP, the journal that had published the original study. The reason given by the editor was *This journal does not publish replication studies, whether successful or unsuccessful* (Aldhous 2011).

When significant results are more likely to get published than non-significant results, the effects that end up in the literature will give a distorted picture of reality (Rosenthal 1979). I will use two simple simulations to illustrate this principle. First, 1000 studies were simulated that compared two samples with independent *t*-tests. The true difference between the populations that the samples were drawn from was set to zero (d=0). In a research environment where all studies are published irrespective of outcome, the mean of the reported effects will accurately describe the true effect. This is illustrated in the "No bias" condition in Figure 1A. However, as illustrated in the "Extreme publication bias" condition in the same figure, the discrepancy between the true effect and the mean of the published effects will be large if only significant results are published.

One thousand new studies were then simulated to illustrate how publication bias distorts the literature when the true effect is different from zero. Like in the previous example, each simulation run created two samples and compared them with an independent *t*-test. This time, the true effect was set to d=0.5 instead of d=0. In a research environment where all positive and negative results are published, the mean of the effects in the literature will accurately describe the true effect. This is illustrated in the "No bias" condition in Figure 1B. By contrast, if only significant effects are published, the average of the published effects will be much larger than the true effect. This scenario is illustrated in the "Extreme publication bias" condition in Figure 1B.

This simple example shows that publication bias not only produce a lot of false positive results (all effects in the "Extreme publication bias" condition in Figure 1A are false positives), but also in an overestimation of effects that are real (almost all effects in both "Extreme publication bias" conditions are inflated).

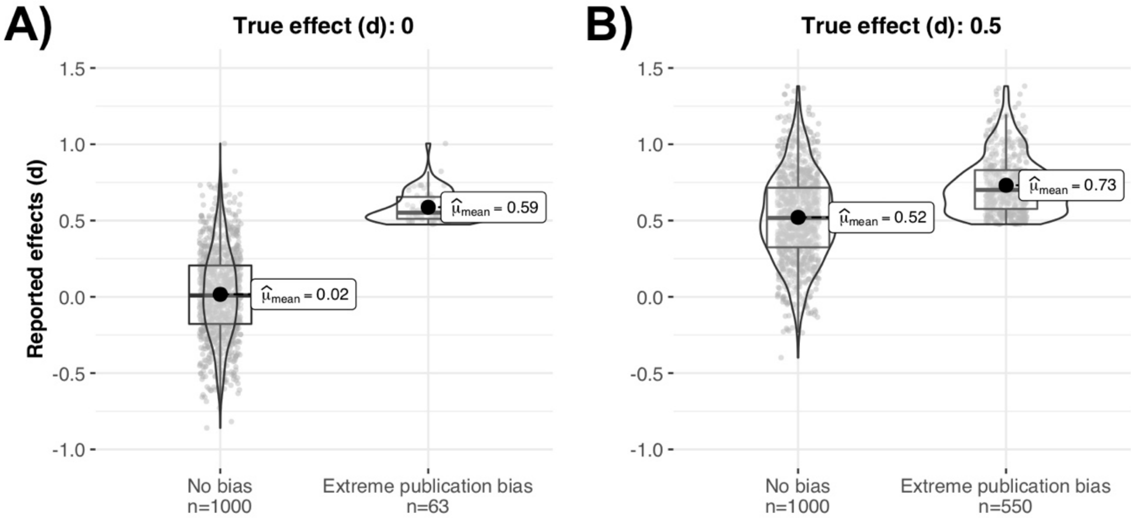

Figure 1A and 1B illustrate the effects of severe publication bias. Each figure displays the distribution of the effect sizes from 1000 simulated independent sample *t*-tests under two different conditions: "No bias" (all 1000 results are reported irrespective of outcome) and "Extreme publication bias" (only significant results are reported). Each gray dot depicts the observed effect from one simulation run, n is the number of published effects per condition, the central horizontal lines illustrate medians, and the black dots means. 1A) The true effect is zero. Without publication bias, the average of the 1000 effects is 0.02. With extreme publication bias, only 63 results out of 1000 are reported and their average effect is 0.59. 1B) The true effect is 0.5 (Cohen's *d*). Without publication bias, the average of the 1000 effects is 0.52. With extreme publication bias, 550 results out of 1000 are reported and their average effect is 0.73.

1.4.2 Questionable research practices

Questionable research practices (QRPs) is another factor that is thought to contribute to the low success rate of replication studies. While publication bias usually refers to situations where negative results are not published, QRPs describe behaviors in the gray zone of what is considered ethical that increase the probability of obtaining positive study outcomes. While there is no generally accepted boundary between these behaviors and outright fraud, the latter is often reserved for the most severe forms of violation of research integrity such as fabricating, falsifying or modifying raw data. Surveys on self-reported research practices indicate that fraud is relatively uncommon (Fanelli 2009). By contrast, QRPs seem to be widespread. In an article from 2012, John, Loewenstein, and Prelec aimed to estimate the prevalence of QRPs by surveying 2155 U.S.-based research psychologists about their engagement in different behaviors that are at risk of producing biased and unreplicable results. As shown in Figure 2, this study indicates that QRPs are, or at least were, common. The most frequently reported self-admitted behaviors were: *Failing to report all dependent measures*, *Collecting more data after seeing whether the results were significant*, *Selectively reporting studies that "worked"*, *Excluding data after looking at the impact of doing so*, and *Claiming to have predicted an unexpected finding*.

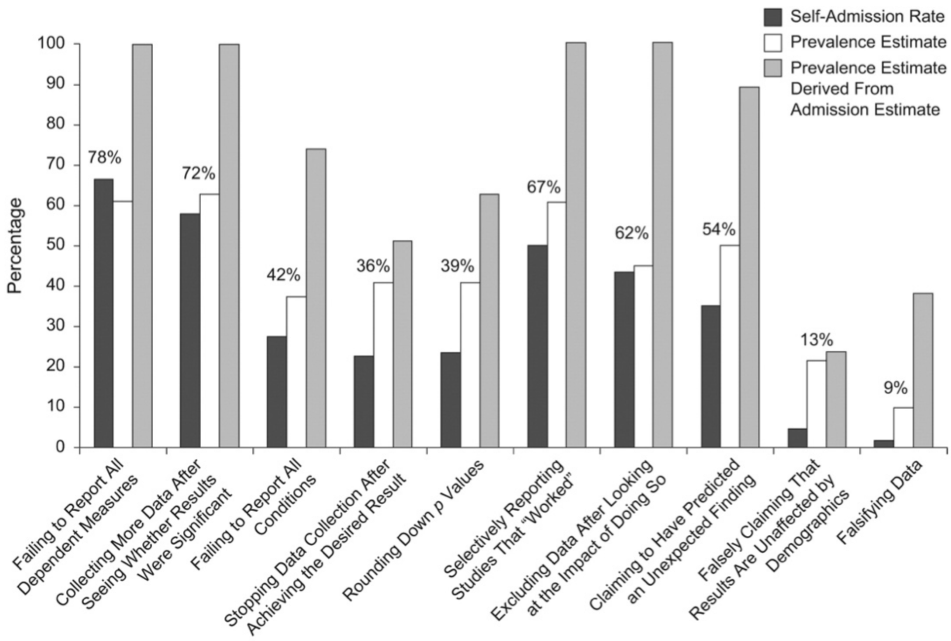

Figure 2 depicts Figure 1 from John et al. (2012). Back bars illustrate the self-reported prevalence of QRPs, white bars the percentage of other psychologists that the respondents thought had engaged in QRPs, and gray bars a prevalence estimate derived from the percentage of the respondents who themselves had engaged in QRPs that would admit to having done so. The numbers above the bars represent the means of these three estimates.

Practices that result in low replicability may not always be perceived as questionable by the researcher (Gelman and Loken 2012). Experimental psychologists are often faced with complex datasets that can be analyzed in multiple ways, especially when the research question is broad and/or unspecific. "Does hunger modulate flavor pleasantness?" is, without a pre-specified analysis plan, an example of an unspecific research question. A main effect of hunger across all rated flavors would obviously suggest that this is the case, however, this is not the only result that could be interpreted as a positive outcome.

Because hunger is associated with a physiological need to restore normal energy levels, an effect on sweet but not on salty flavors could easily be justified post hoc given that sweet foods contain a lot of fast carbohydrates. Moreover, an effect on salty but not on sweet flavors could be interpreted as evidence that sweet flavors are inherently rewarding and therefore less sensitive to contextual manipulations. Hence, there are so far three different results (effect of hunger on all flavors, effect of hunger on sweet flavors, effect of hunger on salty flavors) that could be interpreted as evidence for an amplifying effect of hunger on flavor. Moreover, the researcher must decide which analysis to use, and how to handle covariates, missing data, outliers, etc. Even if the dataset only contains noise, a positive result can be the most likely outcome when there is a high degree of analytical flexibility (Simmons, Nelson, and Simonsohn 2011). Researchers that conduct several tests at random and only report the significant ones are likely aware that what they are doing is questionable. However, researchers that look for visual patterns in the data and then perform a single analysis to test if that pattern is significant may not be aware that such practices also generate a lot of unreplicable results. It is perfectly acceptable to update one's beliefs in light of new data. However, basing analytical decisions on the patterns that might arise in a particular dataset invalidates the usual interpretation of p-values. In the long run, such practices produce a high rate of results that do not generalize to other datasets.

1.4.3 Bias in the flavor literature

To summarize, direct replications have revealed that a large proportion of published results cannot be replicated by independent research teams. Publication bias and questionable research practices (QRPs) are thought to be two primary causes of this problem. Their combined effect is to increase the false-positive rate and inflate effect sizes that reflect true phenomena, which will result in low replicability. It is important to remember that direct replications are still rare (Makel et al. 2012) and that replicability may vary between different subfield of psychology. Because no systematic attempt has been made to estimate the replicability of chemosensory research, the extent to which these problems affect the credibility of the flavor literature remains unknown.

This means that the replicability of the evidence for associative learning may be lower, higher, or similar to 50%, the approximate average of the success rates in large-scale direct replications (15 successes out of 28 [Klein et al. 2018], 10/13 [Klein et al. 2014], 3/10 [Ebersole et al. 2016], 35/97 [Open Science Collaboration 2015], 11/18 [Camerer et al. 2016], 13/21 [Camerer et al. 2018]). One methodological strength of many flavor studies that may benefit replicability is the frequent use of within-participant designs. Such studies may, due to their greater statistical power, generate results that are more stable than results from between-participant designs for any given number of participants. In line with this principle, there is some experimental evidence that replicability is stronger in cognitive (where within-participant designs are the norm) than in social psychology (where between-participant designs are more common). However, replications of studies associated with cognitive and social psychology both seem to produce effect sizes that are much smaller compared with the original effects (Open Science Collaboration 2015). Moreover, the self-admitted rate of QRPs is high among experimental researchers in general, and similar between cognitive and social psychologists (John et al. 2012).

While the best way to assess the credibility of published claims may be to conduct direct replications, there are many things that individual researchers can do to make their results more trustworthy. Methodological standards are rapidly changing to improve different aspects of the scientific process and enable independent researchers to evaluate the work of others. Preregistration and open science practices such as sharing research materials, raw data, and analysis scripts are some examples of new methods that can increase the credibility of quantitative research outcomes (Munafò et al. 2017). While such practices are rare in flavor research, it is important to remember that a large proportion of the studies on associative learning were published in a time when discussions about replicability mostly took place among methodologists. However, based on recent work in this field, it seems to me that flavor researchers have been rather hesitant to adopt these new practices.

In addition to investigating the role of associative learning in flavor perception, a personal goal for me has been the learn about and implement measures to reduce the influence of bias in hypothesis-driven research. I strongly believe that this will be helpful for researchers that wish to build on my work.

2 RESEARCH AIMS

This dissertation contains three studies that in different ways have explored characteristics of flavor that are thought to rely on associative learning. The overarching questions that were addressed in these studies are:

1. Does the degree of congruency between odor and taste modulate: (1) the pleasantness of their combined flavor, and (2) the extent to which the flavor's odor component is referred, or mislocalized, the mouth. **Study I** provides insights into the perceptual and hedonic effects of associations between specific odors and tastes that have already been established thorugh exposure.

2. Does the effect of congruency on flavor pleasantness interact with the hunger state of the perceiver? To optimize food intake, the reward value of flavor varies depending on whether the perceiver is hungry or sated. **Study II** investigates whether the amplifying effect of hunger on pleasantness is stronger for congruent than for incongruent flavors. Such a mechanism would promote consumption of foods that have been determined by previous exposures to be safe and nutritious.

3. Can repeated exposure to unfamiliar odor-taste mixtures create new associations between the olfactory and gustatory modalities? While Study I and Study II investigated hedonic and perceptual effects of already established semantic associations, **Study III** tested whether exposure to unfamiliar odor-taste combinations causally affects characteristics of flavor that are thought to rely on associative learning.

3 METHODS

3.1 TRANSPARENCY

While transparency in itself does not guarantee quality (transparent research can still contain errors, be biased, or provide little to no evidential value), it does create access to the information needed for a proper evaluation of the quality. In response to the realization that many findings cannot be replicated (Open Science Collaboration 2015), researchers have started to adopt open science practices to allow others to scrutinize their work. Such practices include sharing experimental materials, analysis scripts and raw data, using preregistration to limit the effects of analytical flexibility and guarantee that data-independent and data-contingent tests are clearly separated, and to report all results in a way that allows for statistical robustness checks. In addition to facilitating independent evaluations of what has already been published, this development has the potential to increase the efficiency and speed of discovery in the scientific community by letting others reuse data and material that have already been paid for. Study II and Study III were designed based on these principles with one important exception, due to restrictions imposed by the ethical permits, the raw data was not shared.

I was not familiar with open science practices when I conducted Study I, and to be honest, transparency was not one of my top priorities at that time. To compensate, I will provide a disclosure statement to the section that describes Study I to highlight some methodological aspects that I think are important to consider when evaluating those results.

3.2 ETHICAL CONSIDERATIONS

The studies in this dissertation were not associated with any known risks of serious adverse events. All stimuli were made of odorous and gustatory products that are commonly used in the food industry, and all stimuli were safe to consume. At the beginning of each testing session, the experiment leader stressed that participation was voluntary and that the participant was allowed to leave at any time without reprisal. Written informed consent was obtained prior to data collection. Because we did not ask for permission to share the raw data, only the research team had access to the ratings and demographic information about the participants.

All studies were approved by the Regional Ethics Review Board in Stockholm. In addition, Study II and III complied with the ethical principles of the Declaration of Helsinki. Study I compiled with all principles except number 35, which states that *Every research study involving human subjects must be registered in a publicly accessible database before recruitment of the first subject* (World Medical Association 2013).

3.3 PARTICIPANT SAMPLES

The sample sizes of Study I (30) and Study III (60) were set a priori. No formal power calculations were conducted, instead, for each study, the sample size was determined based on available resources and previous studies on similar topics. For Study II, we planned to include 40 participants based on the same considerations. However, this data was collected in early 2020 and due to the onset of the covid-19 pandemic, the final sample size was reduced to 23. With the exception of one individual that did not show up for the second testing session during Study III, results were based on data from all participants. Participants were recruited via a university hosted website where people could sign up for experimental studies. Most were Swedish or European students.

Similar inclusion/exclusion criteria were used for all studies. To be eligible, participants had to be 18-45 years old. Exclusion criteria were tobacco use, current cold or flu symptoms, self-reported taste or smell dysfunction, and less than 11 out of 16 points on an olfactory screening test where 10 or below indicates olfactory dysfunction (Sniffin' Sticks, Hummel, Sekinger, Wolf, Pauli, Kobal, 1997). Because chicken powder was used as stimulus material in Study I and II, no vegetarians and vegans were included in these studies. All participants provided written informed consent and received a small payment on completion. For Study I and Study III, participants were instructed not to eat or drink flavored beverages one hour before the testing sessions to limit potential odor acuity effects. For Study II, a six-hour pre-session fast was required to enhance the effect of the experimental manipulation (intake of a standardized meal).

3.4 STATISTICAL APPROACH

3.4.1 Handling repeated measures data

The datasets from Study I, II, and III contained multiple ratings from each participant. This hierarchical data structure violates the independence assumption of generalized linear models, which increases the risk type I error by producing artificially small *p*-values (Freeberg and Lucas 2009). One way to take the dependencies between the data points into account is to use random effects models (Winter 2013). Let's look at a hierarchical dataset. Figure 3A displays the raw pleasantness ratings from Study II (Study II was chosen because it had the smallest sample size, which makes it easy to visualize each participant separately). The pattern of the boxplots clearly shows that different participants had different baselines, some people liked most flavors while others disliked most flavors.

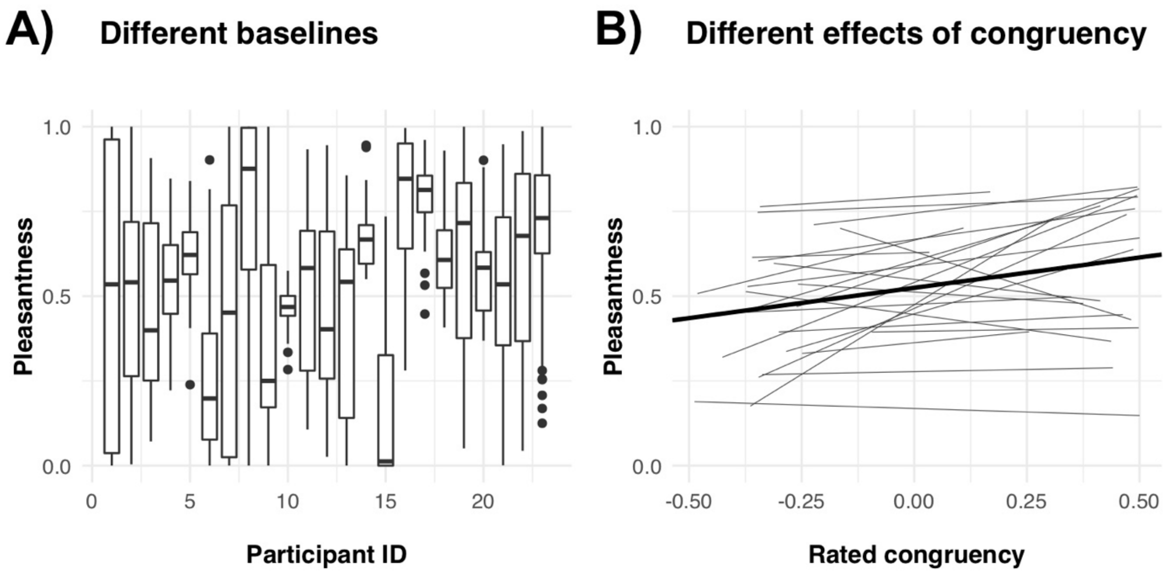

Figure 3A and 3B illustrate a dataset that can be analyzed with models containing by-participant intercepts and slopes for the effect of congruency. 3A: Raw pleasantness ratings plotted separately for each participant. Some participants (e.g. 8, 16, and 17) liked most flavors, while others (e.g. 6, 9, and 15) disliked most flavors. The non-independence between the data points can be accounted for by using a model with random intercepts. 3B: Predicted pleasantness ratings, one linear slope per participant and a thick, black line that depicts the estimated population effect. The effect of congruency varies between individuals. This can be accounted for by using a random slope model.

Mathematically, the dependencies caused by by-participant variation in average pleasantness can be accounted for by assigning each participant to a different intercept. The structure of a random intercept model with Congruency as the only fixed predictor is given by this formula:

$$Pleasantness_{si} = \beta_0 + \beta_1 * Congruency + \mu_{0s} + \varepsilon_{si}$$

Here, $Pleasantness_{si}$ is participant s's pleasantness rating during trial i. β_0 is the grand mean intercept, β_0 the expected increase in pleasantness when congruency increases by one unit, and ε_{si} is the residual (i.e. the deviation between the observed and expected rating for participant s during trial i). μ_{0s} is the random intercept for participant s. This term reflects how much more or less pleasant this particular participant thought the favors were on average.

Just like it is unrealistic to assume that every participant has the same intercept, it may also be unrealistic to assume that the effect of the experimental manipulation is the same for everyone. As shown in Figure 3B, the effect of congruency was not constant across participants. For some people, pleasantness increased strongly across the congruency spectrum indicating that they were sensitive to subtle changes in the experimental manipulation. However, others seemed to like highly congruent and highly incongruent flavors just as much. To handle this variation, we can add a second random effect to the model to allow participants to have different slopes for the effect of congruency. The new formula will look like this:

$$Pleasantness_{si} = \beta_0 + \beta_1 * Congruency + \mu_{0s} + \mu_{1s} * Congruency + \varepsilon_{si}$$

The only difference between this model and the previous one is the inclusion of μ_{1s}, which describes how much stronger or weaker the effect of congruency on pleasantness is for participant s. Importantly, the individual random effects (μ_{0s} and μ_{1s}) are not estimated directly, instead, they are assumed to be drawn from two normal distributions centered at zero. That the distributions peak at zero implies that random effects reflect deviations from the estimated average population effect. The practical implication of this is that the model does not have to estimate two additional parameters per participant (resulting in a total of 46 for this dataset where *n*=23), but rather, it only has to estimate the variances that characterize the distributions from which the individual random effects are drawn. It is this mathematical trick that enables generalization beyond the particular participant sample.

To prevent creating biased standard errors that inflate the type I error rate, it has in general been recommended to use the maximal random effect structure justified by the design during confirmatory hypothesis-testing (Barr et al. 2013). For the studies included in this dissertation, this means including both random intercepts and slopes for any within-participant effects, including interactions. However, complex random effect structures can sometimes prevent models from converging, especially if some of the random effects are not supported by the data. It can then be necessary to simplify the model by removing random slopes, one at the time. To control the type I error rate, methodologists have suggested that the best option is keep the random slope associated with the focus of the confirmatory test for last (Brauer and Curtin 2018). For example, if the hypothesis concerns an interaction between two within-participant predictors, it is advised to remove the random slopes for the main effects before removing the random slopes for the interaction. The analysis plans for Study II and III were designed based on these recommendations. For Study I, we chose to only include random effects that improved model fit.

3.4.2 Interpreting non-significant results

Sensitivity analyses and equivalence tests can be used to make non-significant results more informative. A non-significant effect can be a true negative (if the true effect is zero) or a false negative (if the true effect is different from zero). In any specific case, it is impossible to know for sure which of the two is correct. A limitation of classical tests in frequentist statistics (where the null hypothesis states that the effect is zero) is that they can provide direct support for the existence of positive and negative effects, but not for the absence of effects. This creates an imbalance where results either come out positive (support for the alternative hypothesis) or inconclusive (no support for the null, no support for the alternative). Because many hypothesized effects in Study II and III were not statistically different from zero, we performed additional exploratory analyses to investigate (1) whether we had sufficient power to detect reasonable effect sizes (sensitivity analyses), and (2) whether our data provided direct support for the absence of meaningful effects (equivalence tests). These two methods will be described in the following sections.

3.4.2.1 Sensitivity analyses

Sensitivity analysis can be used post hoc to determine the smallest effect size that a specific test had high power to detect. This approach is based on Monte Carlo simulations (DeBruine and Barr 2021) and involves repeatedly simulating data that is similar to the original dataset in terms of size, structure, and most population parameters. The only exception is the magnitude of the specific effect of interest, which can be set to any value from tiny too large. For a given (assumed) effect size, the sensitivity of the specific test equals the proportion of times that the effect becomes significant. The outcome of a sensitivity analysis can thus reveal how likely it would have been to detect an effect size that is considered reasonable based on observed features in the raw data. If a test has low power to detect meaningful effect sizes, then the conclusion must be that a more powerful design is needed to answer the research question. On the other hand, if a test has excellent power to detect even the smallest effect size of interest, it will seem probable that there is no (meaningful) effect.

3.4.2.2 Equivalence tests

Equivalence tests is a class of analysis methods that can be used to provide direct support for the null hypothesis (Lakens 2017; Lakens, Scheel, and Isager 2018). In frequentist statistics, the procedure starts by specifying an upper and a lower equivalence bound around zero that defines the smallest effect size that is considered theoretically or practically meaningful. Based on the observed data, two separate tests can then assess whether the true effect can be assumed to be (1) significantly smaller than the upper bound, and (2) significantly bigger than the lower bound. If both tests are significant, the conclusion that follows is that the magnitude of the true effect falls within the specified equivalence bound, which means that it is too close to zero to be considered meaningful. It should be noted that it is not possible to test the hypothesis that the true effect is exactly zero, instead, what can be done is to provide evidence that the true effect is too small to be interesting.

4 STUDY I

4.1 SPECIFIC AIMS

The aim of this study was to investigate the role of congruency in two key aspects of flavor: odor referral to the mouth and pleasantness. Odor referral was chosen because it is thought to be part of the reason why humans confuse odor with taste during food consumption, an illusion that may reflect that the unisensory components have been merged into one multisensory object (Spence 2016). Pleasantness was chosen because it regulates the reward value of food. We hypothesized that congruency would have positive effects on both outcomes. This would indicate that repeated exposure to specific odor-taste combinations results in a more unitary and pleasant flavor experience.

4.2 METHODS

Thirty participants were included in this study (M_{age}=26.90, SD_{age}=5.35; 21 women, 9 men). Each participant completed one 90-minute session during which different flavor stimuli were sampled and rated.

Citrus and chicken were chosen as target flavors to represent clearly distinguishable familiar foods. Their odorants and tastants were prepared separately and combined in different ways to create flavor mixtures with varying degrees of congruency. Two pure odorants were used: citrus and chicken. These odorants were approximately isointense and did not evoke somatosensory or gustatory sensations. The pure citrus and chicken served as endpoints on a nine step linear mixture series, progressing from 100% citrus to 100% chicken in steps of 12.5%. Two tastants were used: one that mimicked the taste component of lemonade (sweet/sour) and one that mimicked the taste component of chicken broth (salty/savory). Each tastant was separately mixed with each of the nine odorants. As illustrated in Figure 4, this procedure resulted in one sweet/sour and one salty/savory dilution series that each contained one congruent (citrus odor + sweet/sour taste, chicken odor + salty/savory taste) and one incongruent (citrus odor + salty/savory taste, chicken odor + sweet/sour taste) endpoint.

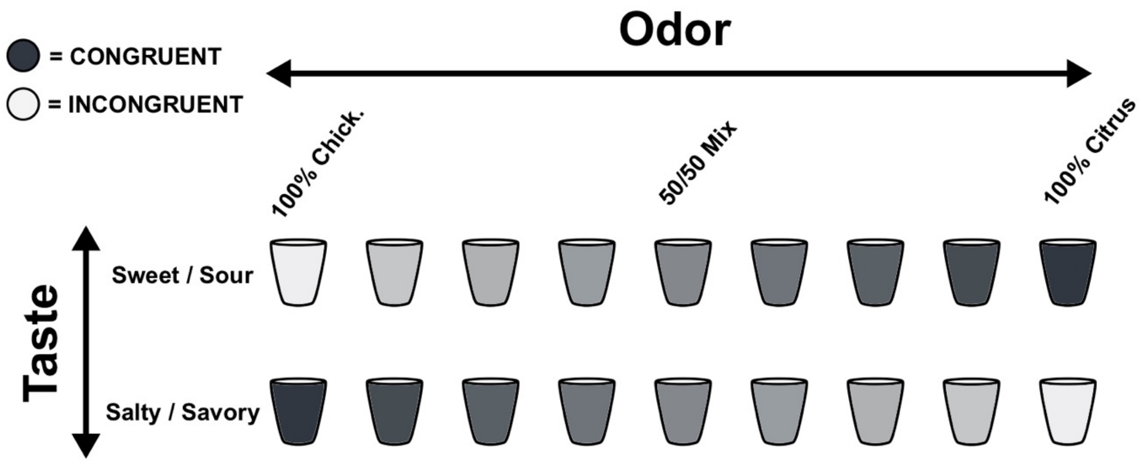

Figure 4. Depiction of the 18 olfactory-gustatory mixtures used in Study I. Odorants are represented on the horizontal axis at the top, each of the 9 vertical flavor-pairs thus have the same odor component. Tastants are represented on the vertical axis at the left. The flavors on the first row have a sweet/sour taste component and the flavors on the second row have a salty/savory taste component. As indicated in the top-left corner, tones of gray illustrate the levels of congruency between odorant and tastant. This figure was adapted from Figure 1 in Fondberg et al. (2018).

Participants were presented with each of the 18 olfactory-gustatory mixtures three times. The nine pure odorants, two pure tastants, and water were each presented once, resulting in a total of 66 trials that were separated into three equally sized blocks with short pauses in between. The presentation order was pseudo-randomized and counterbalanced between participants to minimize order-effects. Before the main experiment, participants were first trained to attend to the localization of the odor component in a vanilla-flavored solution, independently of any taste and somatosensory sensations. To report where they perceived the vanilla odor, participants were shown a cross-sectional drawing of a human head (adapted from Lim and Johnson 2012), which displayed “Oral Cavity”, “Tongue”, “Nose”, and “Throat”, and were asked to select the anatomical location that best matched their experience.

Each trial during the main experiment was initiated by a prompt to taste (but not swallow) one of the solutions. While the participant held the solution in the mouth, two questions appeared consecutively on the screen in random order. One was, “How much do you like the beverage?”, to which the participant answered by clicking on a visual analog scale with the endpoint anchors “Not at all” and “Very much”. The scale consisted of a continuous blue bar positioned across the lower section of the screen. Clicking on the scale logged the response and immediately removed the image. The other question was “In which location(s) do you perceive the odor?”. The cross-sectional drawing of the human head used during training was displayed on the screen for the participant to indicate the localization of the odor. Clicking on a “Continue” button after having selected none, one or several of the locations removed the image. A pause separated the stimulus exposures, during which the participant was instructed to expectorate the solution, rinse with deionized water, and expectorate the rinsing water. Figure 5 depicts the design of the trials.

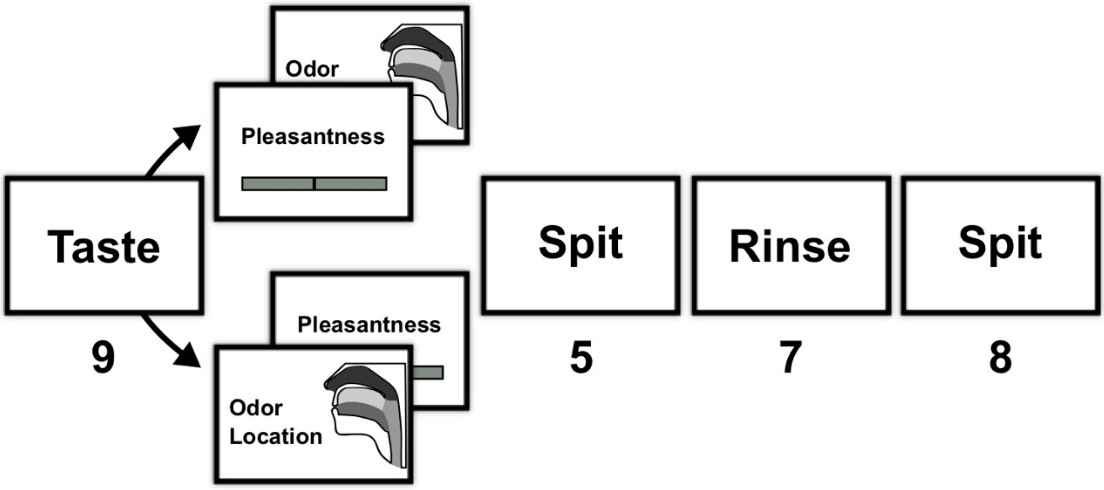

Figure 5. Schematic illustration of the trial design in Study I. During each trial, participants rated the pleasantness of the solution ("How much do you like the beverage?"), and indicated where in the nasal/oral region they perceived the odor component ("In which location(s) do you perceive the odor?"). Numbers indicate presentation duration in seconds. Participants were allowed to take as much time as they needed during the rating tasks.

4.3 RESULTS

As hypothesized, a significant effect of congruency on odor referral to the oral cavity suggested that the more congruent the flavor, the more likely the odor component was to be perceived in the oral cavity ($\chi^2(1)=6.02$, $p=.01$). However, odor referral to the tongue was not significantly affected by congruency ($\chi^2(1)=0.18$, $p=.67$). These results are depicted in Figure 6A.

We then analyzed the pleasantness ratings. In line with our hypothesis, the effect of congruency was highly significant ($\chi^2(1)=17.04$, $p<.001$). Pleasantness increased 9.58 points on the 1-100 scale between the congruent and the incongruent endpoints. By modelling adjusted pleasantness ratings that accounted for differences in odor pleasantness (flavor pleasantness - pleasantness of the flavor's odor component), we found no evidence that the effect of congruency varied between flavors with different taste components (Congruency by Taste component interaction: $\chi^2(1)=0.01$, $p=.94$). In the article (Fondberg et al. 2018), we concluded:

In line with our hypothesis, this indicates that the relationship between pleasantness and congruency was similar for sweet/sour and salty/savory mixtures.

However, a non-significant effect only means that the interaction was not strong enough to be statistically significant. Given the rather wide confidence interval around the observed effect (95% CI=[-0.09, 0.08]), I think that this result should have been interpreted with more caution. Therefore, I reran the analysis separately for sweet/sour and salty/savor flavor and found that as predicted by our initial hypothesis, pleasantness did increase significantly with congruency in both datasets ($\chi^2_{sweet/sour}(1)=5.69$, $p=.02$; $\chi^2_{salty/savory}(1)=5.86$, $p=.02$). These results are depicted in Figure 6B.

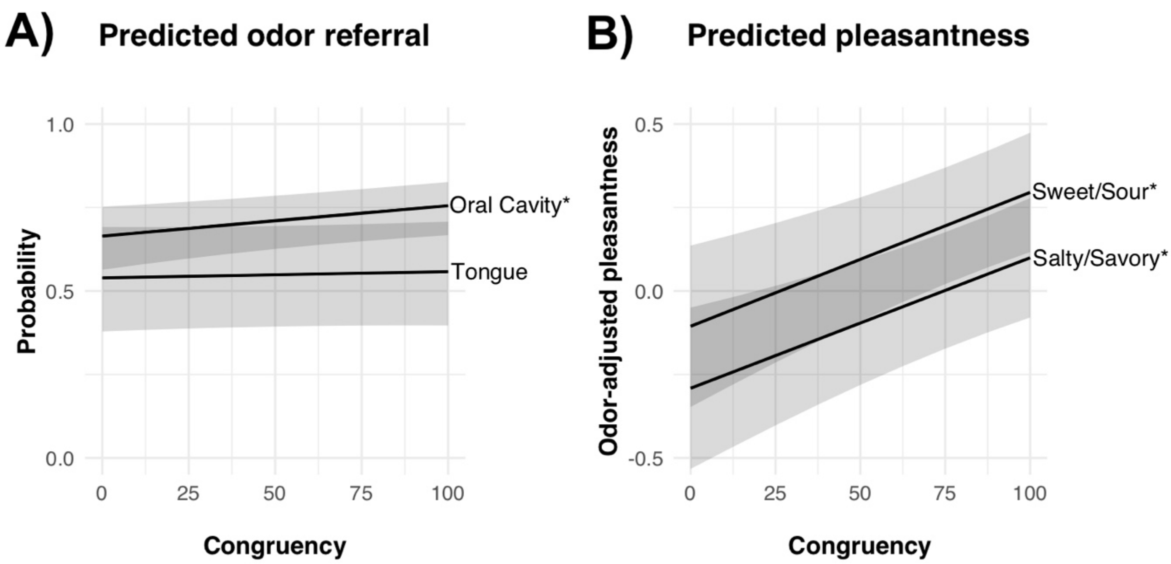

Figure 6A and 6B depicts the results from Study I. Asterisks indicate statistical significance at the 5%-level, 95% confidence intervals are depicted in gray. 6A: The effect of congruency on the probability of odor referral to the oral cavity and tongue. Estimates are derived from models with by-participant random intercepts and a fixed effect of congruency. 6B: Odor-adjusted pleasantness ratings for sweet/sour (top) and salty/savory (bottom) flavors. Odor-adjusted pleasantness was obtained by subtracting the pleasantness ratings of the odor component from its associated flavor pleasantness ratings. These scores were then z-transformed and analyzed with a model that contained random intercepts, random slopes for the effect of congruency, and fixed effects of congruency, taste components and the congruency by taste component interaction.

Finally, we performed a mediation analysis to test whether the effect of congruency on pleasantness was mediated by odor referral to the oral cavity. We found no support for this notion (mediation effect < 0.001, 95% CI: [-0.01, 0.01]).

4.4 DISCLOSURE STATEMENT

All manipulations (congruency) and outcome variables (odor referral, pleasantness) were reported in the article. Before any flavor ratings had been collected, one participant was excluded due to a low score on the olfactory screening test. All data points from the remaining participants were included in the analyses. Before looking at the data, we had hypothesized that pleasantness and odor referral to the mouth would increase linearly with congruency. The mediation analysis was exploratory.

The datasets from Study I and Study II were very similar, both had hierarchical structures and contained congruency scores (based on the mixture composition) and pleasantness ratings. Unlike Study II, Study I was not pre-registered and the analysis plan was therefore developed while the data was being explored. It is difficult to predict if the large number of researcher degrees of freedom that followed from this way of working affected our conclusions. Therefore, to explore the robustness of our results, I retested the hypotheses by using the statistical approach from Study II that had been specified without any knowledge of the research outcomes. The main difference between the approaches was that the models in Study I only included random slopes that improved model fit, while the models in my Study II included the maximal random effect structure that did not cause problems with convergence. The significant effect of congruency on pleasantness was unaffected by this change. However, the effect of congruency on odor referral to the oral cavity was no longer significant (b=0.12, 95% CI=[-0.03, 0.25], $\chi^2(1)$=2.67, p=.10). Moreover, in hindsight, I find it difficult to theoretically motivate the separation between odor referral to the oral cavity, and odor referral to the tongue. We found a significant effect on referral to the former but not to the latter, however, previous studies that have used this approach have found the opposite pattern (Lim et al. 2014; Lim and Johnson 2011, 2012). In an attempt to test for effects on odor referral to the mouth irrespective of exact location in the oral region, I therefore merged the tongue and oral cavity ratings into one outcome variable that was set to 1 when participants localized the odor to any of those anatomical locations and 0 otherwise. A model with by-participant random intercepts, but no slopes (to use the least conservative approach), provided no evidence that referral to the mouth was affected by congruency (b=0.08, 95% CI=[-0.10, 0.25], $\chi^2(1)$=0.71, p=.40).

All of these analyses, including those from the published article, are reasonable and can be justified on theoretical or practical grounds. Nevertheless, when faced with a research question that can be assessed in multiple ways, e.g. by defining odor referral to the mouth as either (1) referral to the oral cavity or (2) referral to the tongue or (3) referral to the oral cavity or tongue, and when the analysis is not data-independent, it is difficult to know which result is most credible. Based on these considerations, my conclusion today is that the evidence from Study I that congruency increases odor referral to the mouth is rather weak. Replications with pre-specified analysis plans are thus highly warranted.

4.5 MAIN CONCLUSION

The significant effect of congruency on pleasantness was the main finding in Study I. This result demonstrated that the stronger the semantic association between the odor and the taste component, the more appetizing the overall flavor sensation becomes. Because semantic associations are learned, the results of Study I thus indicate that repeated exposure has a positive effect on the hedonic dimension of commonly encountered foods. The effect of congruency was present in both sweet/sour and salty/savory flavors, indicating some generalizability across sensory contexts. Moreover, Study I also provides weak support for the notion that congruency affects odor referral to the oral cavity. If this effect is replicable, it would suggest that repeated exposure facilitates the formation of unitary flavor experienced inside the mouth.

5 STUDY II

5.1 TRANSPARENCY

Study II was pre-registered prior to data collection. The registration, analysis scripts, and all research material needed to replicate this study can be found on the Open Science Framework (https://osf.io/n7wyc/). All documents are licensed under CC-By Attribution 4.0 International. We did not apply for permission from the Regional Ethics Review Board to share the raw data, hence, data is only shared on request. All prespecified tests were reported independent of their outcome and analyses that were not pre-registered were labelled exploratory. All data points from the final participant sample were included in the analyses.

5.2 SPECIFIC AIMS

The aim of Study II was to test whether the effect of congruency on flavor pleasantness interacts with the hunger state of the perceiver. While previous studies have suggested that (1) flavor pleasantness increases with congruency (Study I, Fondberg et al. 2018), and (2) flavors pleasantness increases with hunger (Cabanac 1971), the interaction between these two effects had never been explored. Study II was therefore designed to test whether the strength of the amplifying effect of hunger on pleasantness varies across the congruency spectrum. We hypothesized that the positive effect of hunger would be stronger for congruent than for incongruent flavors, which would result in a particularly large increase in pleasantness of foods that are familiar and safe. Finally, in an attempt to explore whether hunger also modulates the perceptual dimension of eating, we aimed to test if flavors are rated as more semantically congruent during hunger than during satiety. We speculated that perceived congruency may act as a mediator that drives hunger-induced increases in pleasantness.

5.3 METHODS

We planned to include 40 participants, but due to the onset of the covid-19 pandemic, the final sample size was reduced to 23 (M_{age}=28.17, SD_{age}=6.31; 15 women, 8 men). Each participant attended two testing sessions that were scheduled to take place 5-15 days apart. The sessions were identical, except that one of them was completed after a fasting period of at least 6 h (Hunger session) and the other one after ingestion of a hummus and feta cheese sandwich with approximately 615 calories (Satiety session).

Study I had defined congruency as the percentage of the odor in the flavor mixture that was assumed to be congruent with the taste. For example, a flavor with a 75% chicken+25% citrus odor and a sweet/sour taste was given a congruency score of 25. Hence, each flavor received a congruency score between 0 (e.g. sweet/sour taste + chicken odor) and 100 (e.g. sweet/sour taste + citrus odor). Because this operationalization does not capture potential non-linear trends in congruency perception, for Study II, we decided to switch to a subjective congruency measure.

Lemonade and chicken broth were chosen as target flavors. They were similar to the target flavors (citrus and chicken) used in Study I. However, based on feedback from the participants in Study I, the mixing recipes of the flavors were slightly adjusted to make the them taste and smell more like real food products. Just as in Study I, the target flavor's unisensory odor and taste components were prepared separately and combined in different ways to create mixtures with varying degrees of semantic overlap. This time, the odor dilution series consisted of 11 odorants that ranged from 100% pure citrus to 100% chicken in steps of 10%. Each of the odorants were mixed with each of the tastants (sweet/sour and salty/savory), which resulted in a total of 22 flavor solutions. The flavors are depicted in Figure 7.

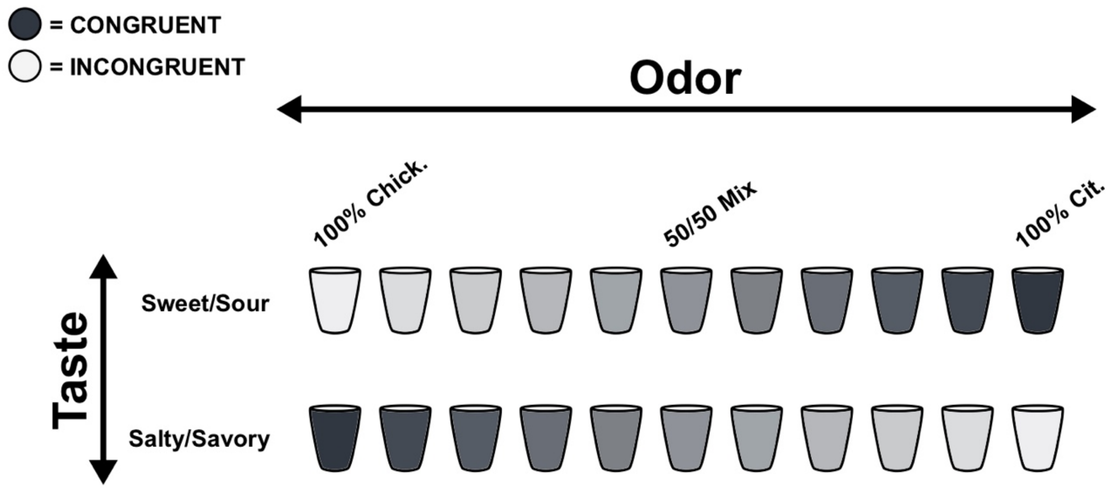

Figure 7. Depiction of the 22 olfactory-gustatory mixtures used in Study II. Odorants are represented on the horizontal axis at the top, each of the 11 vertical flavor-pairs have the same odor component. Tastants are represented on the vertical axis at the left, the flavors on the first row have a sweet/sour taste component and the flavors on the second row have a salty/savory taste component. As indicated in the top-left corner, tones of gray illustrate the degree of congruency (defiend based on the mixture composition, like in Study I) between odorant and tastant. This figure was adapted from Figure 1 in Fondberg et al. (2018).

During the Satiety session, participants were instructed to eat the whole hummus and feta cheese sandwich and drink as much water as they wanted. Based on their order of inclusion, participants were assigned in a 1:1 ratio to attend either the Hunger or the Satiety session first. Both sessions contained two experimental tasks, one that assessed perceived congruency and one that assessed pleasantness. During each task, the 22 flavors, the 11 odorants, the 2 tastants, and plain water were presented once in pseudorandom order.

The congruency task was completed first. Other studies have measured perceived congruency by asking participants to rate the extent to which two stimuli are appropriate for combination in a food product (Schifferstein and Verlegh 1996). We found this operationalization inappropriate for our purposes because: (1) Not all odor-taste pairs that are appropriate for combination in a food product are associated with the same food object. Some unfamiliar combinations (e.g. sea salt and dark chocolate) are unexpectedly appropriate, even when they have rarely been perceived together before and thus lack established associations. (2) Appropriateness is closely related to pleasantness. Most food stimuli that are labelled "appropriate" are also pleasant by definition, and most inappropriate stimuli are unpleasant. We thus suspected that this measure would create an artificially high correlation between "perceived congruency" and pleasantness. To avoid that problem, we conceptualized perceived congruency solely based on the extent to which the combined sensation was perceived as a belonging to either one of two familiar food objects: lemonade or chicken broth. As shown in Figure 8B, participants were instructed to place each stimulus on a scale that ranged from a highly congruent lemonade flavor, via a tick mark in the middle that represented an incongruent 50/50 mix, to a highly congruent chicken broth flavor. The numbers under the scale depict the congruency score (ranging from -0.5 to 0.5) associated with some selected locations.

Participants were first instructed to taste the solutions one at a time and attend to the oral sensation. The congruency scale was then shown on the screen together with the question "How similar to lemonade/chicken broth is this taste?" Clicking on the scale logged the response and immediately removed the image.

The pleasantness task designed to measure pleasantness was completed last. On each trial, participants were instructed to taste a solution and answer the question "How much do you like this taste?". A visual analog scale anchored by "Not at all" and "Very much" was used to collect the ratings. The pleasantness scale in depicted in Figure 8B under the congruency scale.

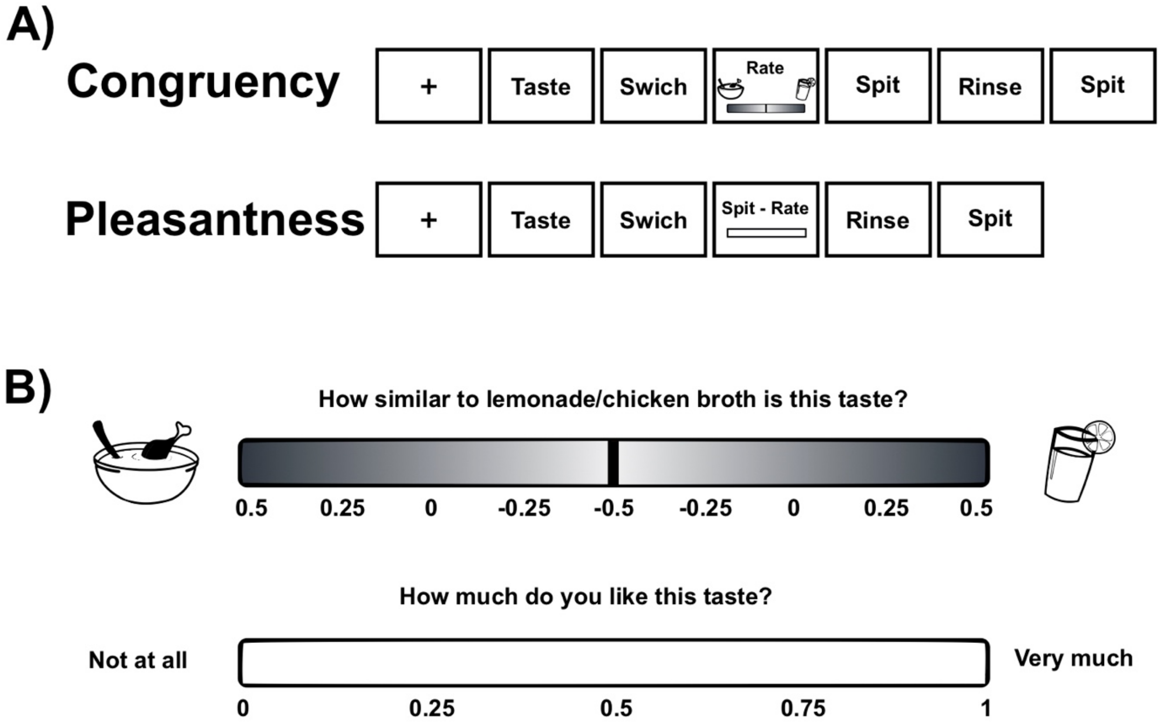

Figure 8A and 8B. Schematic illustration of trial designs and scales used in Study II. 8A: Depiction of the congruency task (top) and the pleasantness task (bottom). Congruency was rated with the solution in the mouth, while pleasantness was rated after the solution had been expectorated. 8B: The scales used to collect congruency (top) and pleasantness (bottom) ratings. Participants were allowed to take as much time as they needed during the rating tasks.

5.4 RESULTS

As expected based on the results from Study I, pleasantness increased linearly with rated congruency ($\chi^2(1)=7.04$, $p=.01$). The predicted increase in pleasantness between the congruent and incongruent endpoints corresponded to 0.17 points on the -0.5:0.5 pleasantness scale (17% of the full rating scale). By contrast, the 0.07 points (7% of the full rating scale) numerical increase in pleasantness between the Satiety session and the Hunger session was not statistically significant ($\chi^2(1)=3.54$, $p=.06$). Similarly, we found no evidence that rated congruency increased with hunger ($\chi^2(1)=2.81$, $p=.09$).

Once these main effects had been assessed, we turned to the primary analysis which focused on the hunger state by rated congruency interaction. Contrary to expectations, we found no evidence that the amplifying effect of hunger increased with congruency ($\chi^2(1)=0.26$, $p=.61$). While the magnitude of this effect was tiny ($b_{\text{interaction}}=-0.03$), the 95% confidence interval was wide [-0.16, 0.10], reflecting the relatively small number of data points. These main results from Study II are depicted in Figure 9.

An exploratory sensitivity analysis was then used to formally investigate the probability of obtaining a significant result, given a wide range of different expected effects and our observed variance parameters and residuals. The results suggested that we only had sufficient power to detect large interactions, corresponding to a slope difference ($b_{\text{cong. during hunger}}$ - $b_{\text{cong. during satiety}}$) of at least 0.2 (20% of the total rating scale).

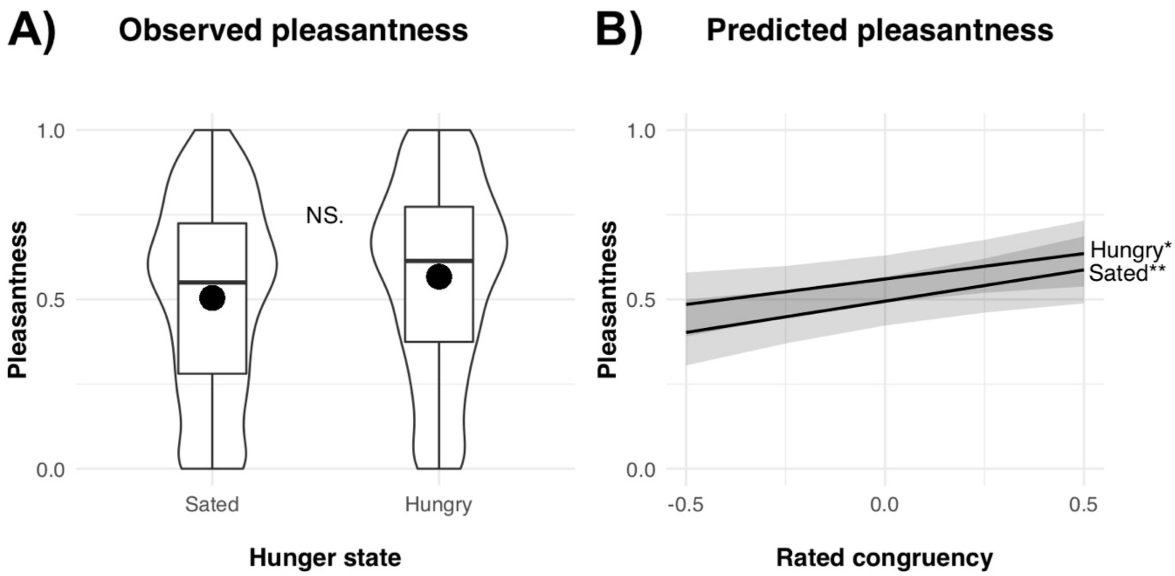

Figure 9A and 9B depicts the results from Study 2. Asterisks indicate significance at the 5% (*) or 1% (**) level, NS=non-significant. 9A: The non-significant difference in pleasantness ratings between hunger states. The central horizontal lines illustrate medians and the black points means. 9B: The significant effects of rated congruency on pleasantness, linear trends and 95% confidence intervals separately for flavors rated during hunger or satiety. The slopes are very similar, which is in line with the non-significant congruency by hunger state interaction.

5.5 MAIN CONCLUSION

The primary hypothesis of Study II was that the amplifying effect of hunger on flavor pleasantness would be stronger for congruent than for incongruent flavors. This perceptual mechanism would promote consumption of familiar foods, which have been determined to be safe due to previous intake. We found no support for this notion. However, an exploratory sensitivity analysis revealed that we only had adequate power to detect large effects. Given the complex nature of flavor perception, most effects in this field, even those that are theoretically relevant, are likely to be small in magnitude. Our non-significant result should thus be interpreted with caution. Moreover, by using congruency ratings as predictor and adhering to a prespecified analysis plan, Study II successfully replicated the positive effect on pleasantness observed in Study I.

6 STUDY III

6.1 TRANSPARENCY

This study was pre-registered before any data had been collected and published with open access in Chemical Senses (https://doi.org/10.1093/chemse/bjab003). All materials (R code, PsychoPy experiments, protocol, stimulus recipes, etc.) needed to replicate this study is available on the Open Science Framework (https://osf.io/dtv8s/) licensed under CC-By Attribution 4.0 International. Like in Study I and II, data is only available on request due to restrictions in the ethics application. We again reported the results from all prespecified tests independently of outcome. No data points were excluded, and all analyses that were not pre-registered were labelled exploratory.

6.2 SPECIFIC AIMS

The aim of Study III was to test whether attributes of flavor that differ between congruent and incongruent odor/taste combinations can be modified directly by exposure. By repeatedly presenting one odor together with taste and another odor alone, we could isolate effects of associative learning within the olfactory/gustatory network from effects of exposure to odors without taste. A screening of the literature revealed four attributes that differ between congruent and incongruent flavors: flavor pleasantness, odor referral to the mouth, odor intensity enhancement by taste, and odor sweetness. Our specific hypothesis was that these attributes would be more affected by exposure to combined odor-taste mixtures than by exposure to pure odors. While Study I and Study II explored potential effects of semantic associations that had already been encoded in memory, Study III was designed to experimentally induce such associations and assess their combined effect on perceptual and hedonic aspects of food perception.

6.3 METHODS

Sixty participants were included in this study (M_{age}=27.31, SD_{age}=5.05; 40 women, 20 men), all of which attended two testing sessions (one pre-exposure session and one post-exposure session) of approximately 75 min that were spaced 5-11 days apart. Between the sessions, participants chewed two types of gums: one that was flavored with an odorant, and one that was flavored with a second odorant and sucrose.

Basil and orange flower were used as target odors. During the testing sessions, the odorants were presented both orthonasally (sniffed from bottles when assessing odor sweetness) and retronasally (sipped from medicine cups when assessing the other outcomes). To produce flavor stimuli, sucrose was added to each of the two odorants at a concentration that perceptually mimicked moderately sweet drinks.

Unsweetened and sweetened chewing gums flavored with the target odors were used to present the stimuli between the testing sessions. Four different flavor categories of gum were prepared, two that only contained odor (pure basil and pure orange flower), and two that contained both odor and taste (basil-sucrose and orange flower-sucrose). At the end of the first session, participants were in alternating order either given 15 sweetened basil gums and 15 unsweetened orange flower gums, or 15 unsweetened basil gums and 15 sweetened orange flower gums. During the next five days, subsequently referred to as the exposure phase, participants chewed six chewing gums per day: two before breakfast, two before lunch, and two before dinner. One of the two chewing gums was always sweetened, and the other unsweetened. Participants were in daily contact with the experiment leader to ensure that no gums were skipped.

The pre and post-exposure sessions each contained three short experiments (Figure 10). The first experiment assessed changes in orthonasal odor sweetness between the two testing sessions. We only expected the perceived sweetness to increase for odors that had been paired with sucrose during the exposure phase. The second experiment assessed changes in retronasal odor pleasantness. Our hypothesis was that rated pleasantness would increase for all flavors due to the effect of mere-exposure. Moreover, we expected the increase in pleasantness of sweet-paired odors to be larger than the increase of unpaired odors, but only for participants that liked sweet taste to begin with. The third experiment assessed changes in odor intensity enhancement by taste, and odor referral to the mouth. We hypothesized that mere-exposure to odors without taste would not affect these phenomena, but that sweet-pairing would result in both increased odor intensity enhancement and increased odor referral.

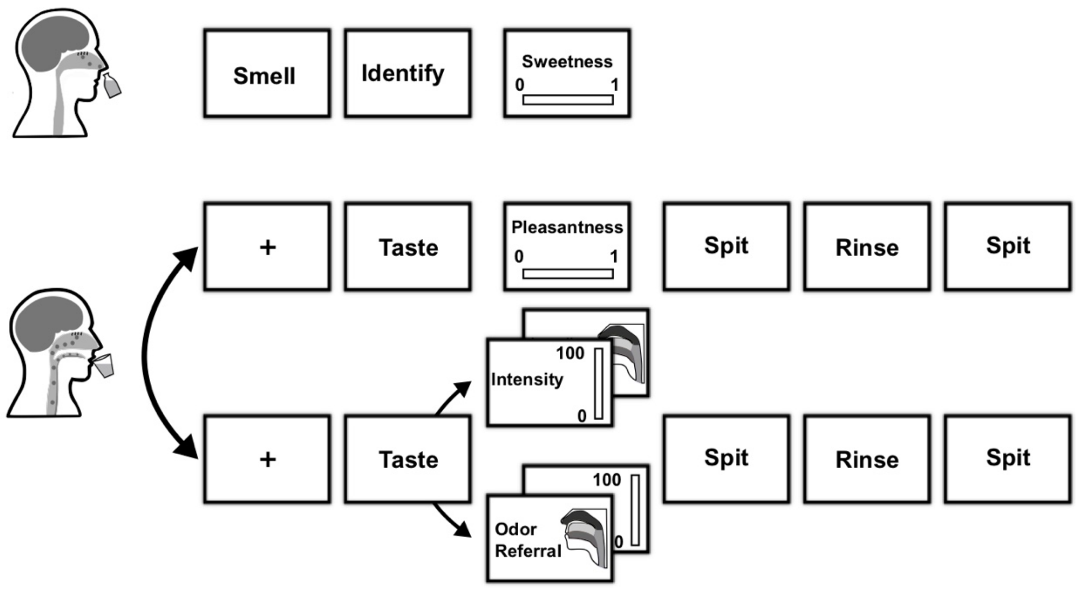

Figure 10. Trial design of the three experiments in Study III. The first row depicts the experiment that assessed orthonasal odor sweetness, the second row shows the experiment that assessed odor pleasantness, and the third row shows the experiment that assessed odor referral to the mouth and odor intensity enhancement by taste. To avoid order effects, the sequence of the latter two experiments was alternated between participants, and intensity and odor referral were assessed in random order. This figure was adapted from Figure 1B in Fondberg, Lundström, and Seubert (2021).

6.4 RESULTS

For odor sweetness, we focused on the Session (pre-exposure, post-exposure) by Condition (sweet-paired odor, unpaired odor) interaction to test whether exposure to odors in combination with sweet taste had resulted in a greater perceptual change than exposure to odors without taste. The non-significant result ($\chi^2(1)=0.94$, $p=.33$) was surprising given that, compared to previous research that has reported significant effects, our participants were exposed to the flavors a large number of times to ensure that the experimental manipulation would be strong, and our participant sample was large enough to detect effect sizes of the magnitude reported in previous studies. An exploratory equivalence test confirmed that the increase in odor sweetness of sweet-paired odors was either absent, or at least significantly smaller than previously thought (equivalence bound = ±10% of the rating scale; upper bound $t(59)=-2.47$, $p=.01$, lower bound $t(59)=5.29$, $p<.001$).

We then analyzed odor pleasantness. As hypothesized, a small but significant main effect of session suggested that the odors were more positively evaluated after exposure ($\chi^2(1)=6.24$, $p=.01$). The increase corresponded to ~4% of the total rating scale and did not vary significantly depending on whether the odor had been exposed with or without sucrose (Session by Condition interaction: $\chi^2(1)=1.60$, $p=.21$). That the increase in pleasantness did not seem to vary between sweet-paired and unpaired odors when analyzing all participants was expected given that not everyone likes sweet taste. To assess whether sweet-pairing had resulted in a larger exposure effect among sweet-likers, we tested whether the (non-significant) Session by Condition term was modulated by the degree to which participants liked sweet taste. We found no evidence that this was the case ($\chi^2(1)=0.29$, $p=.59$). Because this result was not in line with the literature, an additional analysis was conducted that assessed the same question in a slightly different way. Specifically, we tested whether the observed increases in pleasantness (post-exposure - pre-exposure) correlated with the participants' sweet-liking scores. Again, we found no evidence for such an effect ($r(58)=-0.15$, $p=.25$). Results from our study thus suggest that the potential effects of associative learning on odor pleasantness might be weaker or more fragile than previously thought.

Similarly, we found no evidence for effects of associative learning on odor intensity enhancement by taste ($\chi^2(1)=0.28$, $p=.60$), or on odor referral to the mouth (Oral cavity: $\chi^2(1)=0.82$, $p=.37$; Tongue: $\chi^2(1)=0.94$, $p=.33$). This was the first time these outcomes were assessed longitudinally in a learning study. Researchers with an interest in these phenomena should design experiments that are powered to detect small effects.

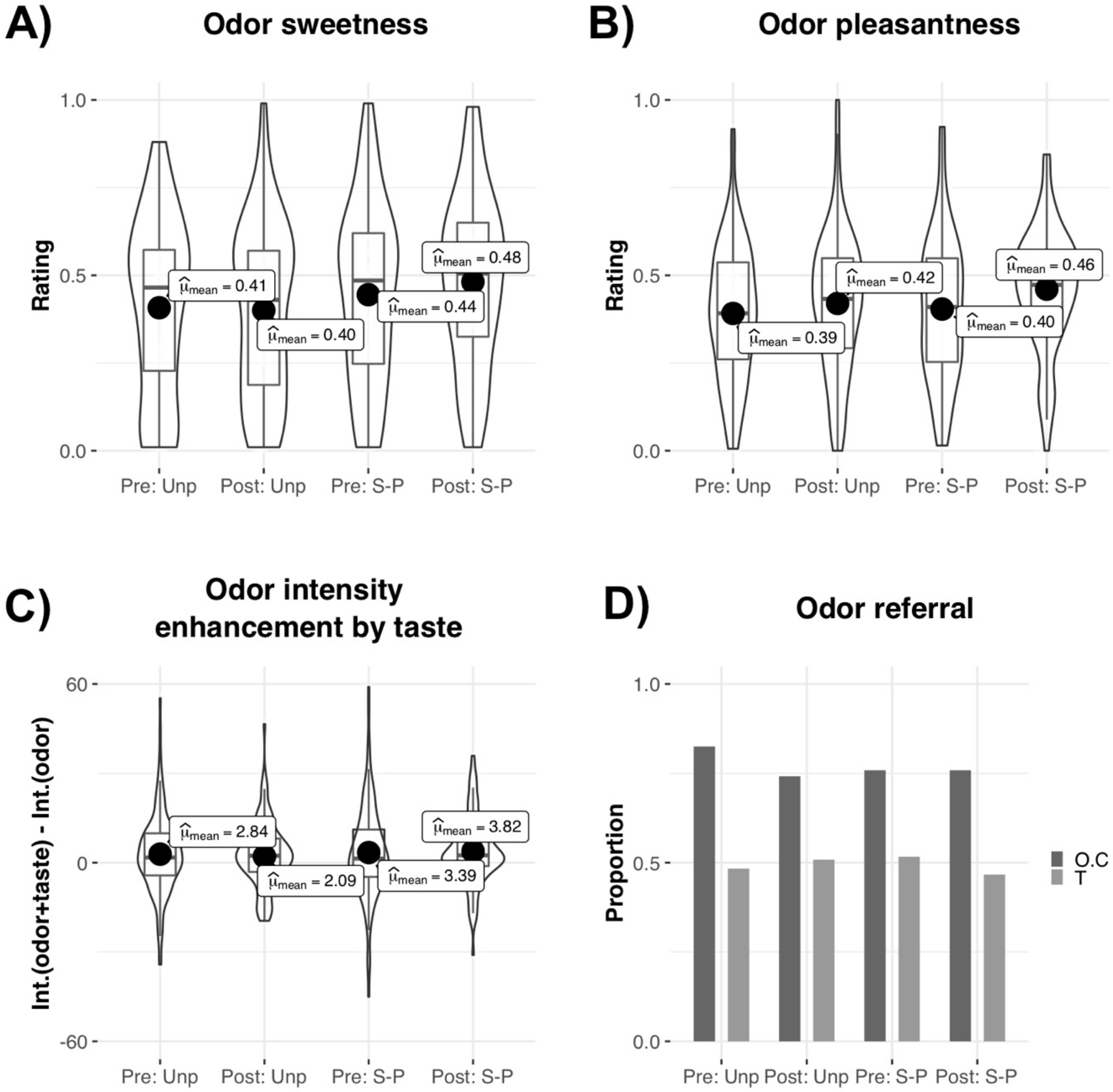

Figure 11 depicts the four outcomes of Study III. Ratings are shown separately for each session (pre-exposure [Pre], post-exposure [Post]) by condition (unpaired odor [Unp], sweet-paired odor [S-P]) context. In each boxplot, the black point displays the mean and the horizontal line the median. The two graphs at the top depict rated odor sweetness (left, 0.01=”not sweet at all”, 1=”extremely sweet”) and rated odor pleasantness (right, 0.01=”not pleasant at all”, 1=”extremely pleasant”). Odor intensity enhancement (bottom left) was calculated by subtracting the intensity ratings (1=”barely detectable”, 100=”strongest imaginable”) of the pure odor solutions, from the intensity ratings of the same odor solutions presented with sucrose. Odor referral (bottom right) shows the proportions of trials that the flavor’s odor component was localized to the mouth (O.C=oral cavity, T=tongue). This figure was adapted from Figure 1B in Fondberg, Lundström, and Seubert (2021).

6.5 MAIN CONCLUSION

In Study III, we tested whether repeated exposure to relatively unfamiliar odor-taste mixtures could induce associative learning within the olfactory-gustatory network. Despite its relatively large sample size, powerful exposure protocol, multiple outcome measures, and liberal alpha-level (5% with no correction for multiple comparisons), Study III failed to provide any evidence of associative learning. Exploratory equivalence tests further suggested that the hypothesized effects were either absent, or at least smaller than in previous studies. One possibility is that this discrepancy is explained by differences between Study III's protocol and previous protocols that have been used to assess associative learning. For example, while other researchers have used flavored solutions to present the stimuli during the exposure phase, we chose to use chewing gums instead so that participants could be exposed to the odors (and sweet taste) several times without having to visit the lab. Another possible explanation for the null results of Study III is that associative learning effects are weaker and less robust than previously thought. A high-powered, direct replication of a study with positive results would be useful to confirm that associative learning effects reliably can be demonstrated in experimental settings.

7 GENERAL DISCUSSION

7.1 SUMMARY OF RESULTS

When we eat, synchronized activation of the olfactory and gustatory modalities results in the creation of holistically perceived flavors that appear to arise from inside the mouth. The observation that frequently encountered odor-taste combinations are processed and perceived differently (Small et al. 2004) than unfamiliar odor-taste combinations suggests that associative learning modulates this binding process. Investigating how interactions within the olfactory-gustatory network are affected by repeated exposure has great potential to advance our understanding of how food preferences develop over time. In this dissertation, I have presented three lab experiments that in different ways have assessed this question.

Study I and Study II explored perceptual and hedonic effects of learning that has already taken place. Specifically, these studies investigated how central features of the flavor percept change as a function of a graded manipulation of congruency. These studies were based on the assumption that odors and tastes that are associated with the same food often will be perceived together, which will generate synchronized activation of the olfactory and gustatory modalities. Presenting the same food over and over again will eventually create a stable, semantic link between the odor and taste components which will be indicated by increased perceived congruency. Investigating how flavor perception changes across the congruency spectrum thus provides an indirect way of obtaining information on how the olfactory-gustatory network is affected by learning.

To be able to present flavors with varying degrees of congruency to the participants, an odor dilution series was first obtained by using two familiar food odors to create different mixtures that ranged from 100% citrus to 100% chicken in steps of ~10%. Each of these odorants were then mixed with two different tastants, one sweet/sour that mimicked the taste of soft drinks, and one salty/savory that mimicked the taste of chicken broth. The final stimulus set thus contained two supposedly congruent flavors: citrus odor+sweet/sour taste (lemonade) and chicken odor+salty/savory taste (chicken broth), two supposedly incongruent flavors: chicken odor+sweet/sour taste and citrus odor+salty/savory taste, and several intermediate mixtures.

In Study I, congruency was operationalized as the proportion of the odor in the odor mixture that was associated with the same food as the taste it was presented with. The results revealed a positive, linear-like effect of congruency on flavor pleasantness. This finding is in line with results from a previous study where highly congruent flavors were rated as more pleasant than highly incongruent flavors (Small et al. 2004). Moreover, Study I provided some evidence that congruency also increases the probability that a flavor´s odor component will be perceived in the oral cavity, but not on the tongue. We interpreted this finding as an indication that the extent to which odors and tastes are merged into flavor depends on how often the combination has been perceived in the past. However, given that the effect on odor referral was not robust to subtle changes in the analytical approach and that previous studies in general have reported referral to the tongue but not to the oral cavity (Lim et al. 2014; Lim and Johnson 2011, 2012), I am no longer convinced that this result reflects a true effect. A careful theoretical and experimental investigation of the validity of a distinction between odor referral to the oral cavity and to the tongue is highly warranted. After that, a strictly controlled pre-registered study with high power to detect small effects could then provide a replicable estimate for the potential effect.

By using a similar stimulus set and adhering to a pre-registered analysis plan, Study II replicated and corroborated the positive effect of congruency on pleasantness from Study I with a new participant sample. This time, congruency was operationalized as the degree to which an odor-tastes mixture resembled a familiar flavor with strongly associated unisensory components. The advantage of this approach was that it could capture potential non-linearities between the ingredients in a flavor solution and the degree to which the odor and taste in combination were perceived as a semantically coherent food object. Study I and Study II provide strong evidence that congruent flavors are more appetizing than incongruent flavors. Importantly, this effect seems to be of substantial magnitude, which indicates that congruency has practical relevance for the eating process.

Turning to the primary hypothesis of Study II, we found no evidence that the amplifying effect of congruency on pleasantness interacts with the hunger state of the perceiver. Specifically, we expected that hunger would have a positive hedonic effect on pleasantness that would be stronger for congruent than for incongruent flavors. However, the observed difference in pleasantness between hunger and satiety was not significant, and we found no evidence that this (non-significant) effect varied with congruency. This observation is consistent with the notion that *if* hunger promotes consumption of foods by making their flavors more pleasant, then familiar and novel foods will be affected to the same extent. With that said, a non-significant result does not in itself provide evidence that the true effect is zero. An exploratory sensitivity analysis revealed that our design only had sufficient power to detect large effects. Given the complex nature of flavor perception and the large number of factors that affect pleasantness, small interactions may seem more plausible than large ones. While the results from Study II do suggest that strong interactions between hunger state and congruency are unlikely to exist, replication studies with high power to detect small effects or test for equivalence are needed for conclusive results.

The aim of Study III was to test whether repeated exposure to relatively unfamiliar odor-taste mixtures is accompanied by measurable changes in how their flavors are perceived. Instead of exploring potential effects of already established associations, this study investigated whether such associations could be experimentally induced by exposure. Like in Study II, we pre-registered the analysis plan to limit researcher degrees of freedom. A screening of the literature revealed four perceptual phenomena that have been suggested to be affected by exposure to odor and (sweet) taste in combination: flavor pleasantness (Schifferstein and Verlegh 1996; Small et al. 2004), odor referral to the mouth (Lim and Johnson 2011, 2012), odor intensity enhancement by taste (the ability of tastes to enhance the intensity of congruent odors (Lim et al. 2014)), and odor sweetness (Stevenson et al. 1995). After a five-day exposure phase where one odor was perceived together with sweet taste and another odor was perceived alone as a control condition, associative learning was not demonstrated for any of the outcomes. Exploratory equivalence tests revealed that the effects were either absent or smaller than in previous studies.

Because previous studies on associative learning have relied on small sample sizes and non-transparent analytical procedures that cannot be guaranteed to have been developed independently of the data at hand, a transparently conducted replication of a study with strong results is needed to confirm that associative learning can reliably be observed in experimental settings. Ideally, such studies should be conducted by independent research teams. Based on the available evidence, I do not think one can conclude that associative learning is an important factor in flavor perception in humans. I do, however, think that there are reasons to be optimistic about the potential of this field. While some aspects of food perception seem to be partially innate (Newcomb, Xia, and Reed 2012), there is little doubt that environmental exposure also is highly important. The question is not whether learning affects flavor, but to what extent learning effects are mediated by associations between the olfactory and gustatory modalities. The observation that most odors that are often perceived with sugar are both pleasant and sweet-smelling suggests that some taste qualities actually can get transferred to the odor via repeated exposure. However, whether experimental conditions can be identified that reliably produce perceptual or hedonic changes after a relatively short exposure phase remains to be determined. A good starting point would be to conduct a direct replication of a successful experiment on odor sweetness, which, based on previous research, seems to be the most promising target.

7.2 FURTHER EXPLORATION OF THE RELATIONSHIP BETWEEN CONGRUENCY AND PLEASANTNESS

The main finding of this dissertation is without doubt the amplifying effect of odor-taste congruency on flavor pleasantness. This effect was first revealed in Study I and then replicated in Study II with a pre-registered analysis. Based on the highly significant result from the linear model in the first study, we argued that the observed effect was linear. However, that an effect is significant actually says very little about the shape of the relationship. Linear and non-linear trends can both give rise to significant results when fitted to linear models and as shown in Figure 5 in that article, visual inspection of the pleasantness ratings actually indicated that a sigmoid-like function may have generated a better model fit.

However, the sample size of Study I may have been too small to test the shape of the relationship in a meaningful way. But because Study II used almost exactly the same stimulus set, procedure, and scale, I am now in a much better position to address this question. There is no theoretical reason to expect any specific shape so instead of comparing different prespecified models to each other, I simply plotted the average pleasantness ratings (+ 95% CI) separately for all measured levels of congruency. Each dataset was first mean centered to remove baseline differences between the studies. Due to the repeated measures design, the congruency levels contained several ratings from each participant. To not violate the independence assumption when calculating the confidence intervals, these ratings were first averaged. This resulted in a dataset with flavor ratings from 53 individuals. Congruency was defined as the percent of odor in the chicken/lemon odor mixture that semantically matched the taste (just like in Study I).

Figure 12 depicts the results. None of the 95% confidence intervals exclude the best fitting linear slope and the positive and negative deviations from linearity are evenly distributed across the congruency spectrum. This exploratory finding needs to be replicated by confirmatory research, however, based on the available evidence, the best guess today is that the amplifying effect of congruency on flavor pleasantness actually is linear. A linear relationship would indicate that minor differences between what is perceived (a specific odor-taste mixture that arises from the mouth) and what is expected (a perfect prototype of the encountered food) only have small, negative effects on the reward value of foods. Although congruent odor-taste mixtures indeed would be most pleasant, slightly less congruent mixtures would not be automatically rejected. Such flexibility would enable the perceiver to appreciate a wide range of flavors, which likely is an important ability given that most flavor categories are quite heterogeneous.

This result may also have implications for the interpretation of the results in Study III, which provided no evidence for associative learning effects. Throughout this dissertation, I have conceptualized congruency as the extent to which an odor and a taste are associated with the same food object. Congruent sensations are attributes of the same external object, while incongruent sensations are attributes of different objects. It is this idea that links Study I and II to Study III, in which exposure to relatively unfamiliar flavors was used to create new associations. These associations were expected to increase the perceived congruency of the unisensory components, and modify the experience of the resulting flavor. The specific odor-taste pairs were chosen to be neither as congruent as the citrus odor+sweet/sour taste mixture, nor as incongruent as the chicken odor+sweet/sour taste mixture. Importantly, the congruent combinations in Study I and II had likely been perceived thousands of times by most participants. Given that the difference in pleasantness between the moderately congruent and the highly congruent mixtures was no more than 10 points on the 1:100 pleasantness scale, experimental designs with 10-20 exposures may only be expected to generate effects that are much smaller than that. However, it is of course possible that other perceptual features, e.g. odor sweetness, display larger effects that are easier to reliably detect in experimental settings.

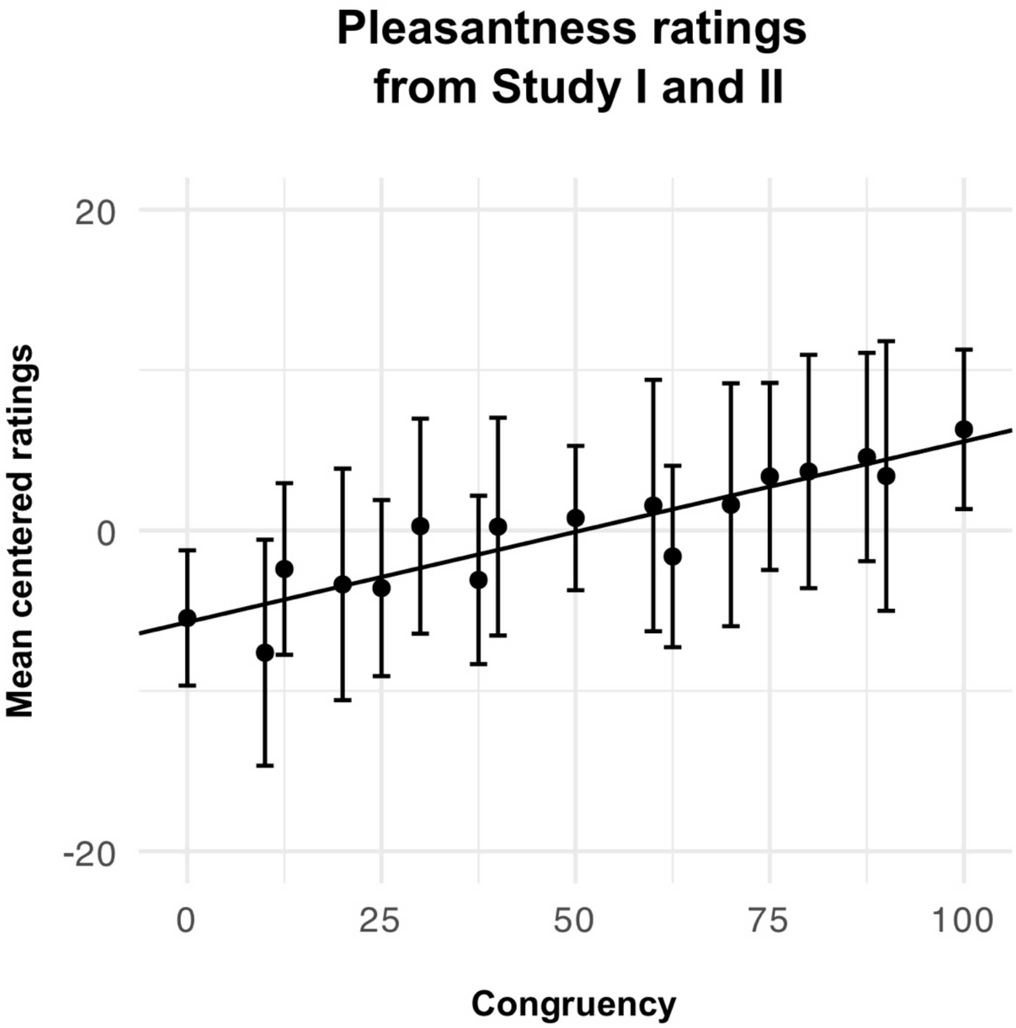

Figure 12 depicts the pleasantness ratings from Study I and Study II separate for each level of congruency. To account for the baseline difference between the studies, both datasets have been mean centered. Error-bars represent 95% confidence intervals and the black line is the best fitting linear slope. The relationship between congruence and pleasantness appears to be linear.

8 FUTURE DIRECTIONS

Few studies on flavor perception are pre-registered and there is great variability in the analysis pipelines that have been used to investigate key questions. Moreover, even though typical sample sizes are small, often around 15-50 participants, null result studies are rare. The extent to which this is a consequence of questionable research practices and publication bias is not known. However, even if these problems are less severe than in other subfields of experimental psychology, I definitely see room for improvement.

The bright side is that things are starting to change. The replication crisis has brought about many changes in the standards by which research is being conducted and evaluated. So far, topics like replicability and transparency have mostly been discussed by social psychologists and meta researchers. However, I think that the time has come for us who investigate other aspects of human nature to also be slightly more skeptical about commonly accepted truths in our fields. In addition, to enhance the credibility of the published literature, we also need to change the way that new evidence is produced. I will in the next couple of paragraphs outline some of the most promising developments that are currently taking place to make research claims more credible.

8.1 OPEN SCIENCE AND PREREGISTRATION

It is not possible to demonstrate quality without transparency. Therefore, researchers need to share the material that was used to arrive at any given conclusion. Scrutinizing the work of others should be seen as a service to the field that enhances the self-correcting feature of the scientific method by sorting out claims that are not supported by the evidence. Sharing research data (in a privacy preserving manner) is equally important. Open data not only enables quality checks, but also allow others to reuse the data to answer new research questions. This practice has great potential to accelerate scientific progress and increase the return of invested funds.

Being transparent about when ideas were generated is another central component of a transparent workflow. When done right, preregistration helps both the researcher and the readers to separate outcomes that result from prediction from outcomes that result from postdiction (Nosek et al. 2018). A prediction is a hypothesis that is generated from the existing literature, intuition, dreams, everyday observations of natural phenomena, computational modelling, or almost any other source of inspiration. The only exception is the data that is going to be used to test the hypothesis. Data-independent analyses that test predictions generate interpretable *p*-values that correspond directly to the strength of the evidence for the hypothesized association. Although testing predictions with new data is vital for separating true claims from false claims, generating new hypotheses based on existing data, e.g. postdiction, is also necessary for scientific progress. However, complex datasets will almost always contain “significant” associations, some of which reflect noise and others of which reflect true effects. Interpreting unexpected findings in the same way as findings that support predictions will inevitably result in an unacceptably high type I error rate, which in turn decreases the field´s replicability. Instead, results that are data-contingent should be considered hypothesis-generating, rather than hypothesis-testing. In general, such results need to be confirmed prospectively with new data before they can be assumed to reflect true phenomena.

Without access to a detailed preregistration, it is impossible for others to know whether the described analyses were conducted in a hypothesis-generating or a hypothesis-testing mode of research. Methodologists have therefore argued that it should be required for researchers that want to make confirmatory claims to register both their hypotheses and an analysis plan in advance (Wagenmakers et al. 2012). I definitely agree and hope that in a few years, studies without a high-quality preregistration will be given much less evidential value than today.

8.2 DIRECT REPLICATIONS

Given that bias is known to affect the trustworthiness of the published literature, there is an urgent need for direct replications of key experiments. While many researchers still hesitate to conduct studies that are designed solely to confirm the work of others, this type of studies are necessary for science to be self-correcting. After a decade of crisis or revolution, depending on one's perspective (Vazire 2018), it has become clear that psychology needs to expand its focus from producing "novel" results to also testing the reliability of results that have already been published. Almost all effects in the flavor literature are described in a way that implies generalizability beyond the exact experimental conditions and participant sample of the particular study. This implies that other researchers should be able to detect it when using similar methods and large samples. Conducting replications that closely resemble the original is thus necessary to separate reliable results that deserve further investigation from unreliable results that are best to be treated as noise (Simons 2014).

Unfortunately, direct replications in the flavor literature are rare, a problem that is definitely not unique to this subfield within psychology (Makel, Plucker, and Hegarty 2012). By contrast, conceptual replications, i.e. studies that test the same theoretical idea by using different methods, are fairly common. This type of research design allows investigators to test the generalizability of effects whose reliability have already been established. The problem that arises when conducting many conceptual, but no direct, replications is that it takes a long time before researchers begin to question false claims. While a successful conceptual replication is almost always interpreted as evidence for the theory in the original study, a failed conceptual replication is much more difficult to interpret (even when it is adequately powered). Null results from such studies could be caused by methodological moderators, however, null results are also to be expected when the original result is a false positive. Study III is an example of a conceptual replication that probably would have been more informative if it had been designed as a direct replication. The observed null-results might reflect that associative learning effects are either absent, or at least not as strong as previously thought.

However, the results might equally well be explained by some methodological aspect of our research design.

The best way to determine which explanation is correct is to conduct an independent, high-powered, pre-registered, transparent, direct replication of a study with strong, positive results. Importantly, this should be a priority before we start to explore potential moderators, mediators, or neural or behavioral correlates.

8.3 SLOW SCIENCE

Obtaining a better understanding of the perceptual mechanisms that regulate our eating experiences should be a prioritized task in chemosensory research. If done right, this field can address urgent questions such as: *Which cognitive and perceptual factors contribute to healthy and pathological eating patterns? Can the reward value of food be manipulated by focused interventions?* and *How does our ability to acquire new food preferences change throughout the life course, and can this ability be enhanced?* To produce results that are not only reliable, but also more useful, I think that flavor researchers need to focus on fewer topics to allow for a greater in-depth examination of selected questions. Identifying the most promising topics and then investing enough resources to get a firm understanding of the construct itself and the factors that regulate it, not only in lab-setting, may require a different way of doing research.

The cornerstones of any large-scale research program should be transparency and equal appreciation of positive and negative results. Initially, the research may be carried out in an exploratory fashion (Scheel, Tiokhin, et al. 2020). Trying out different ways of modelling and measuring the construct, identifying and exploring auxiliary assumptions, experimenting with different stimulus material and procedures, and qualitatively evaluating promising outcomes are some examples of meaningful activities at this stage. Importantly, these results should be honestly reported as part of a hypothesis-generating process. When the research team has a firm understanding of what they want to test and how they want to test it, the next step is to identify experimental conditions that reliably produce the hypothesized result. This means conducting high-powered, direct replications of successful experiments until a method has been identified that almost always "works".

After that, different aspects of the experimental design can be changed in a controlled manner to determine how generalizable the finding is across different conditions. For example, while the effect of congruency on pleasantness in Study I and II does seem to be replicable, we do not know whether it is flavor specific (we used citrus and chicken odor and sweet/sour and salty/savory taste in both studies) or reflects a more general perceptual phenomenon. After all, if congruency only affects the pleasantness of meaty and citrus-like flavors, it may not require further investigation. This is a general problem in chemosensory research that probably cannot be solved without structural reforms. Selecting, preparing, presenting, and storing odorants and tastants require a lot of time and effort, which likely explains why most studies in this field (including my own) only have used one or two flavors. However, generalizing from such a small sample to the entire population of flavors may be just as problematic as generalizing from one or two participants to all healthy adults (Yarkoni 2020).

When the research team has identified effects that reliably can be produced in strictly controlled lab-experiments using different stimuli, the next step is to test the effects in real-life settings. Investigations in ecologically valid environments with real foods will provide valuable information on the practical relevance of the findings.

To reduce bias towards positive results, short term successes (defined as significant results published in high-impact journals) need to be decoupled from the researchers' competitiveness on the academic labor market. This would require major reforms of the incentive structures in academia that may seem unlikely today, however, I do think that the growing realization that a lot of research efforts are wasted due to suboptimal practices will eventually make funders and other important stakeholders recognize the need for change.

9 CONCLUDING THOUGHTS

In this dissertation, my primary aim has been to highlight the most important results from Study I, II and III, together with some methodological aspects that I think are particularly important to consider when evaluating research in this field. Hopefully, my commitment to open science practices has reduced the influence of bias and produced materials and results that will be useful to other researchers in this field. The results presented in this dissertation have not revolutionized our understanding of the perceptual factors that regulate eating behavior, however, I do think that I have made a small but significant contribution to the field by first of all revealing and replicating a linear-like effect of congruency on flavor pleasantness, and second, to present null-results that highlight the need for independent and rigorous replications of associative learning effects.

10 ACKNOWLEDGEMENTS

First and foremost, I want to thank **Sam** for putting up with my ridiculously long hours, keeping my feet on the ground, and making me believe that I could do this. Without your support, I would never have made it this far.

I want to thank my main supervisor **Janina Seubert** for taking me on as PhD-student and guiding me through this process, and my co-supervisors **Johan N. Lundström** and **Mats J. Olsson** for always having an open door when I needed someone to talk to.

I am lucky enough to have some brilliant friends who also happen to be extremely generous with their time. Thank you **Marin Asperholm**, **Jens Fust**, **Malin Freidle**, **Moa Peter**, **Danja Porada**, **Martin Schaefer**, and **Kristin Öster** for providing valuable feedback on earlier drafts of this dissertation.

I also want to express my heartfelt gratitude to my mentors, **Rickard Carlsson** and **Petter Gustavsson**. I have great respect for your knowledge and integrity.

Last but not least, I want to mention the people in the open science communities I am a part of. Members of **ReproducibiliTea** and **Open Science Community-Sweden**, thank you for making me believe that the future of science is bright.

11 REFERENCES

Aldhous, P. (2011, May 5). Journal Rejects Studies Contradicting Precognition. *New Scientist*. Retrieved from https://www.newscientist.com/

Baeyens, F., Crombez, G., Hendrickx, H., and Eelen, P. (1995). Parameters of Human Evaluative Flavor-Flavor Conditioning. *Learning and Motivation*, *26*(2), 141–60. doi:10.1016/0023-9690(95)90002-0

Barr, D. J., Levy, R., Scheepers, C., and Tily. H. J. (2013). Random Effects Structure for Confirmatory Hypothesis Testing: Keep It Maximal. *Journal of Memory and Language, 68*(3), 255-278. doi:10.1016/j.jml.2012.11.001

Bear, M. F., Connors, B. W., and Paradiso, M. A. (2007). *Neuroscience: Exploring the Brain*, 3rd Ed. Philadelphia, PA, US: Lippincott Williams & Wilkins Publishers.

Beauchamp, G. K., and Mennella, J. A. (2011). Flavor Perception in Human Infants: Development and Functional Significance. *Digestion, 83*(Suppl. 1), 1–6. doi:10.1159/000323397

Bem, D. J. (2011). Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect. *Journal of Personality and Social Psychology, 100*(3), 407–25. doi:10.1037/a0021524

van den Bosch, I., van Delft, J. M., de Wijk, R. A., de Graaf, C, and Boesveldt, S. (2015). Learning to (Dis)like: The Effect of Evaluative Conditioning with Tastes and Faces on Odor Valence Assessed by Implicit and Explicit Measurements. *Physiology & Behavior, 151*, 478–84. doi:10.1016/j.physbeh.2015.08.017

Brauer, M., and Curtin, J. J. (2018). Linear Mixed-Effects Models and the Analysis of Nonindependent Data: A Unified Framework to Analyze Categorical and Continuous Independent Variables That Vary within-Subjects and/or within-Items. *Psychological Methods, 23*(3), 389–411. doi:10.1037/met0000159

Cabanac, M. (1971). Physiological role of pleasure. *Science, 173*(4002), 1103-1107. doi:10.1126/science.173.4002.1103

Camerer, C. F., Dreber, A., Forsell, E., Ho, T. H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., and Wu, H. (2016). Evaluating Replicability of Laboratory Experiments in Economics. *Science, 351*(6280), 1433–36. doi:10.1126/science.aaf0918

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., Isaksson, S., Manfredi, D., Rose, J., Wagenmakers, E. J., and Wu, H. (2018). Evaluating the Replicability of Social Science Experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour, 2*(9), 637–44. doi:10.1038/s41562-018-0399-z

Chandrashekar, J., Hoon, M. A., Ryba, N. J., and Zuker, C. S. (2006). The Receptors and Cells for Mammalian Taste. *Nature, 444*(7117), 288–94. doi:10.1038/nature05401

Chen, X., Gabitto, M., Peng, Y., Ryba, N. J., and Zuker, C. S. (2011). A Gustotopic Map of Taste Qualities in the Mammalian Brain. *Science, 333*(6047), 1262–66. doi:10.1126/science.1204076

Chikazoe, J., Lee, D. H., Kriegeskorte, N., and Anderson, A. K. (2019). Distinct representations of basic taste qualities in human gustatory cortex. *Nature communications, 10*(1), 1-8. doi:10.1038/s41467-019-08857-z

Crouzet, S. M., Busch, N. A., and Ohla, K. (2015). Taste Quality Decoding Parallels Taste Sensations. *Current Biology, 25*(7), 890–96. doi:10.1016/j.cub.2015.01.057

DeBruine, L. M., and Barr, D. J. (2021). Understanding Mixed Effects Models through Data Simulation. *Advances in Methods and Practices in Psychological Science, 4*(1). doi:10.1177/2515245920965119

Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., ... and Nosek, B. A. (2016). Many Labs 3: Evaluating Participant Pool Quality across the Academic Semester via Replication. *Journal of Experimental Social Psychology 67*, 68–82. doi:10.1016/j.jesp.2015.10.012

Fanelli, D. (2009). How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data. *PLOS ONE, 4*(5):e5738. doi:10.1371/journal.pone.0005738

Fanelli, D. (2010). 'Positive' Results Increase Down the Hierarchy of the Sciences. *PLOS ONE, 5*(4):e10068. doi:10.1371/journal.pone.0010068

Firestein, S. (2001). How the Olfactory System Makes Sense of Scents. *Nature, 413*(6852):211–18. doi:10.1038/35093026

Fondberg, R., Lundström, J. N., Blöchl, M., Olsson, M. J., and Seubert, J. (2018). Multisensory Flavor Perception: The Relationship between Congruency, Pleasantness, and Odor Referral to the Mouth. *Appetite, 125*, 244–52. doi:10.1016/j.appet.2018.02.012

Fondberg, R., Lundström, J. N., and Seubert, J. (2021). Odor-Taste Interactions in Food Perception: Exposure Protocol Shows No Effects of Associative Learning. *Chemical Senses, 46*. doi:10.1093/chemse/bjab003

Frank, R. A., and Byram, J. (1988). Taste–Smell Interactions Are Tastant and Odorant Dependent. *Chemical Senses, 13*(3), 445–55. doi:10.1093/chemse/13.3.445

Freeberg, T. M., and Lucas, J. R. (2009). Pseudoreplication Is (Still) a Problem. *Journal of Comparative Psychology, 123*(4), 450–51. doi:10.1037/a0017031

Freiherr, J. (2017). Cortical Olfactory Processing. In A. Buettner (Eds.), *Springer Handbook of Odor*, Springer Handbooks, Cham, Switzerland: Springer International Publishing

Fujimaru, T, and Lim, J. (2013). Effects of Stimulus Intensity on Odor Enhancement by Taste. *Chemosensory Perception, 6*(1), 1–7. doi:10.1007/s12078-013-9140-5

Gelman, A, and Loken, E. (2012). The Garden of Forking Paths: Why Multiple Comparisons Can Be a Problem, Even When There Is No 'Fishing Expedition' or 'p-Hacking' and the Research Hypothesis Was Posited Ahead of Time. *Department of Statistics, Columbia University*, 348:17

Gerkin, R. C., and Castro, J. B. (2015). The Number of Olfactory Stimuli That Humans Can Discriminate Is Still Unknown. *ELife, 4.* doi:10.7554/eLife.08127

Gibson, J. J. 1966. *The Senses Considered as Perceptual Systems*. Boston, MA, US: Houghton Mifflin

Gottfried, J. A. (2010). Central Mechanisms of Odour Object Perception. *Nature Reviews Neuroscience, 11*(9), 628–41, doi:10.1038/nrn2883

Green, B. G., Nachtigal, D., Hammond, S, and Lim, J. (2012). Enhancement of Retronasal Odors by Taste. *Chemical Senses, 37*(1), 77–86. doi:10.1093/chemse/bjr068

Hausner, H., Olsen, A., and Møller, P. (2012). Mere exposure and flavour–flavour learning increase 2–3 year-old children's acceptance of a novel vegetable. *Appetite, 58*(3), 1152-59. doi:10.1016/j.appet.2012.03.009

Hollingworth, H. L., and A. T. Poffenberger Jr. (1917). *The Sense of Taste*. New York, NY, US: Moffat & Yard doi:10.1037/10825-000

Hummel, T., and Heilmann, S. (2008). Olfactory Event-Related Potentials in Response to Ortho- and Retronasal Stimulation with Odors Related or Unrelated to Foods. *International Dairy Journal, 18*(8), 874–78. doi:10.1016/j.idairyj.2007.10.010

Hummel, T., Heilmann, S., Landis, B. N., Reden, J., Frasnelli, J., Small, D. M., and Gerber, J. (2006). Perceptual Differences between Chemical Stimuli Presented through the Ortho- or Retronasal Route. *Flavour and Fragrance Journal, 21*(1), 42–47. doi:10.1002/ffj.1700

Hummel, T., Sekinger, B., Wolf, S. R., Pauli, E., and Kobal, G. (1997). 'Sniffin'sticks': olfactory performance assessed by the combined testing of odor identification, odor discrimination and olfactory threshold. *Chemical senses, 22*(1), 39-52. doi:10.1093/chemse/22.1.39

Iwata, S., Yoshida, R., and Ninomiya, Y. (2014). Taste Transductions in Taste Receptor Cells: Basic Tastes and Moreover. *Current Pharmaceutical Design, 20*(16), 2684–92. doi:10.2174/13816128113199990575

John, L. K., Loewenstein, G., and Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices with Incentives for Truth Telling. *Psychological Science, 23*(5), 524–32. doi:10.1177/0956797611430953

Johnson, B. A., and Leon, M. (2007). Chemotopic Odorant Coding in a Mammalian Olfactory System. *The Journal of Comparative Neurology, 503*(1), 1–34. doi:10.1002/cne.21396

Kepes, S., and McDaniel, M. A. (2015). The Validity of Conscientiousness Is Overestimated in the Prediction of Job Performance. *PLOS One, 10*(10):e0141468. doi:10.1371/journal.pone.0141468

Kinnamon, S. C., and Finger, T. E. (2019). Recent advances in taste transduction and signaling. *F1000Research, 8*. doi:10.12688/f1000research.21099.1

Klein, R. A., Ratliff, K. A., Vianello, M., Adams Jr, R. B., Bahník, Š., Bernstein, M. J., ... and Nosek, B. A. (2014). Investigating Variation in Replicability. *Social Psychology, 45*(3), 142–52. doi:10.1027/1864-9335/a000178

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams Jr, R. B., Alper, S., ... & Sowden, W. (2018). Many Labs 2: Investigating Variation in Replicability Across Samples and Settings. *Advances in Methods and Practices in Psychological Science, 1*(4), 443–90. doi:10.1177/2515245918810225

Labbe, D., and Martin, N. (2009). Impact of Novel Olfactory Stimuli at Supra and Subthreshold Concentrations on the Perceived Sweetness of Sucrose after Associative Learning. *Chemical Senses, 34*(8), 645–51. doi:10.1093/chemse/bjp045

Lakens, D. (2017). Equivalence Tests: A Practical Primer for t Tests, Correlations, and Meta-Analyses. *Social Psychological and Personality Science, 8*(4), 355–62. doi:10.1177/1948550617697177

Lakens, D., Scheel, A. M., and Isager, P. M. (2018). Equivalence Testing for Psychological Research: A Tutorial. *Advances in Methods and Practices in Psychological Science, 1*(2), 259–69. doi:10.1177/2515245918770963

Lim, J., Fujimaru, T., and Linscott, T. D. (2014). The Role of Congruency in Taste–Odor Interactions. *Food Quality and Preference*, *35,* 5–13. doi:10.1016/j.foodqual.2013.12.002

Lim, J., and Johnson, M. B. (2011). Potential Mechanisms of Retronasal Odor Referral to the Mouth. *Chemical Senses, 36*(3), 283–89. doi:10.1093/chemse/bjq125

Lim, J, and Johnson, M. B. (2012). The Role of Congruency in Retronasal Odor Referral to the Mouth. *Chemical Senses, 37*(6), 515–22. doi:10.1093/chemse/bjs003

Mainland, J. D., Keller, A., Li, Y. R., Zhou, T., Trimmer, C., Snyder, L. L., ... and Matsunami, H. (2014). The Missense of Smell: Functional Variability in the Human Odorant Receptor Repertoire. *Nature Neuroscience 17*(1),114–20. doi:10.1038/nn.3598

Makel, M. C., Plucker, J. A, and Hegarty, B. (2012). Replications in Psychology Research: How Often Do They Really Occur? *Perspectives on Psychological Science, 7*(6), 537–42. doi:10.1177/1745691612460688

McCabe, C., and Rolls, E. T. (2007). Umami: A Delicious Flavor Formed by Convergence of Taste and Olfactory Pathways in the Human Brain. *The European Journal of Neuroscience, 25*(6), 1855–64. doi:10.1111/j.1460-9568.2007.05445.x

Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Du Sert, N. P., ... & Ioannidis, J. P. (2017). A Manifesto for Reproducible Science. *Nature Human Behaviour, 1*(1), 1–9. doi:10.1038/s41562-016-0021

Mura Paroche, M., Caton, S. J., Vereijken, C. M., Weenen, H., Houston-Price, C. (2017). How Infants and Young Children Learn About Food: A Systematic Review. *Frontiers in Psychology, 8*, 1046. doi:10.3389/fpsyg.2017.01046

Newcomb, R, D., Xia, M. B., and Reed, D. R. (2012). Heritable Differences in Chemosensory Ability among Humans. *Flavour, 1*(1), 1-9. doi:10.1186/2044-7248-1-9

Nosek, B. A., Ebersole, C. R., DeHaven, A. C, and Mellor, D. T. (2018). The Preregistration Revolution. *Proceedings of the National Academy of Sciences, 115*(11), 2600–2606. doi:10.1073/pnas.1708274114

Obels, P., Lakens, D., Coles, N. A., Gottfried, J., and Green, S. A. (2020). Analysis of Open Data and Computational Reproducibility in Registered Reports in Psychology. *Advances in Methods and Practices in Psychological Science, 3*(2), 229–37. doi:10.1177/2515245920918872

Open Science Collaboration. (2015). Estimating the Reproducibility of Psychological Science. *Science, 349*(6251). doi:10.1126/science.aac4716

Porcu, E., Benz, K. M., Ball, F., Tempelmann, C., Hanke, M., and Noesselt, T. (2020). Macroscopic information-based taste representations in insular cortex are shaped by stimulus concentration. *Proceedings of the National Academy of Sciences, 117*(13), 7409-7417. doi:10.1073/pnas.1916329117

Prescott, J. (1999). Flavour as a Psychological Construct: Implications for Perceiving and Measuring the Sensory Qualities of Foods. *Food Quality and Preference, 10*(4), 349–56. doi:10.1016/S0950-3293(98)00048-2

Prescott, J. (2012). Chemosensory Learning and Flavour: Perception, Preference and Intake. *Physiology & Behavior, 107*(4), 553–59. doi:10.1016/j.physbeh.2012.04.008

Prescott, J., Johnstone, V., and Francis, J. (2004). Odor-Taste Interactions: Effects of Attentional Strategies during Exposure. *Chemical Senses, 29*(4), 331–40. doi:10.1093/chemse/bjh036

Prescott, J., and Murphy, S. (2009). Inhibition of Evaluative and Perceptual Odour–Taste Learning by Attention to the Stimulus Elements. *Quarterly Journal of Experimental Psychology, 62*(11), 2133–40. doi:10.1080/17470210903031169

Ritchie, S. J., Wiseman, R., and French, C. C. (2012). Failing the Future: Three Unsuccessful Attempts to Replicate Bem's 'Retroactive Facilitation of Recall' Effect. *PLOS ONE, 7*(3):e33423. doi:10.1371/journal.pone.0033423

Roper, S. D., and Chaudhari, N. (2017). Taste Buds: Cells, Signals and Synapses. *Nature Reviews Neuroscience, 18*(8):485–97. doi:10.1038/nrn.2017.68

Rosenthal, R. (1979). The File Drawer Problem and Tolerance for Null Results. *Psychological Bulletin, 86*(3), 638–41. doi:10.1037/0033-2909.86.3.638

Rozin, P. (1982). 'Taste-Smell Confusions' and the Duality of the Olfactory Sense. *Perception & Psychophysics, 31*(4), 397–401. doi:10.3758/BF03202667

Scheel, A. M., Schijen, M., and Lakens, D. (2020). An Excess of Positive Results: Comparing the Standard Psychology Literature with Registered Reports. *Advances in Methods and Practices in Psychological Science, 4*(2), doi:10.1177/25152459211007467

Scheel, A. M., Tiokhin, L., Isager, P. M., and Lakens, D. (2020). Why Hypothesis Testers Should Spend Less Time Testing Hypotheses. *Perspectives on Psychological Science.* doi:10.1177/1745691620966795

Schifferstein, H. N., and Verlegh, P. W. (1996). The Role of Congruency and Pleasantness in Odor-Induced Taste Enhancement. *Acta Psychologica, 94*(1), 87–105. doi:10.1016/0001-6918(95)00040-2

Schuenke, M., and Schulte, E. (2011). *Head and Neuroanatomy*. Stuttgart, Germany: Thieme

Seubert, J., Ohla, K., Yokomukai, Y., Kellermann, T., and Lundström, J. N. (2015). Superadditive Opercular Activation to Food Flavor Is Mediated by Enhanced Temporal and Limbic Coupling. *Human Brain Mapping, 36*(5), 1662–76. doi:10.1002/hbm.22728

Simmons, J. P., Nelson, L. D., and Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science, 22*(11), 1359–66. doi:10.1177/0956797611417632

Simons, D. J. (2014). The Value of Direct Replication. *Perspectives on Psychological Science, 9*(1), 76–80. doi:10.1177/1745691613514755

Small, D. M. (2012). Flavor Is in the Brain. *Physiology & Behavior, 107*(4), 540–52. doi:10.1016/j.physbeh.2012.04.011

Small, D. M., Voss, J., Mak, Y. E., Simmons, K, B., Parrish, T., and Gitelman, D. (2004). Experience-Dependent Neural Integration of Taste and Smell in the Human Brain. *Journal of Neurophysiology, 92*(3), 1892–1903. doi:10.1152/jn.00050.2004

Small, D. M., Zald, D. H., Jones-Gotman, M., Zatorre, R. J., Pardo, J. V., Frey, S., and Petrides, M. (1999). Human Cortical Gustatory Areas: A Review of Functional Neuroimaging Data. *NeuroReport, 10*(1), 7–13. doi:10.1097/00001756-199901180-00002

Smith, D. V., and Boughter Jr, J. D. (2009). *Taste: Vertebrate Central Pathways*. in Encyclopedia of Neuroscience. Cambridge, Massachusetts, US: Academic Press.

Spence, C. (2015a). Just How Much of What We Taste Derives from the Sense of Smell? *Flavour, 4*(1), 30. doi:10.1186/s13411-015-0040-2

Spence, C. (2015b). Multisensory Flavor Perception. *Cell, 161*(1), 24–35. doi:10.1016/j.cell.2015.03.007

Spence, C. (2016). Oral Referral: On the Mislocalization of Odours to the Mouth. *Food Quality and Preference, 50*:117–28. doi:10.1016/j.foodqual.2016.02.006

Spence, C., and Piqueras-Fiszman, B. (2014). *The Perfect Meal: The Multisensory Science of Food and Dining.* New York, US: John Wiley & Sons

Stein, B. E., and Stanford, T. R. (2008). Multisensory Integration: Current Issues from the Perspective of the Single Neuron. *Nature Reviews Neuroscience, 9*(4), 255–66. doi:10.1038/nrn2331

Sterling, T. D., Rosenbaum, W. L., and Weinkam, J. J. (1995). Publication Decisions Revisited: The Effect of the Outcome of Statistical Tests on the Decision to Publish and Vice Versa. *The American Statistician, 49*(1):108–12. doi:10.1080/00031305.1995.10476125

Sterling, T. D. (1959). Publication Decisions and Their Possible Effects on Inferences Drawn from Tests of Significance-Or Vice Versa. *Journal of the American Statistical Association, 54*(285), 30–34. doi:10.2307/2282137

Stevenson, R. J. (2009). *The Psychology of Flavour*. Oxford, UK: Oxford University Press

Stevenson, R. J. (2014). Flavor Binding: Its Nature and Cause. *Psychological Bulletin, 140*(2), 487–510. doi:10.1037/a0033473

Stevenson, R. J., Boakes, R. A., and Prescott, J. (1998). Changes in Odor Sweetness Resulting from Implicit Learning of a Simultaneous Odor-Sweetness Association: An Example of Learned Synesthesia. *Learning and Motivation, 29*(2), 113–32. doi:10.1006/lmot.1998.0996

Stevenson, R. J., and Case, T. I. (2003). Preexposure to the Stimulus Elements, but Not Training to Detect Them, Retards Human Odour-Taste Learning. *Behavioural Processes, 61*(1–2), 13–25. doi: 10.1016/s0376-6357(02)00166-3

Stevenson, R. J., Prescott, J., and Boakes, R. A. (1995). The Acquisition of Taste Properties by Odors. *Learning and Motivation, 26*(4), 433–55. doi:10.1016/S0023-9690(05)80006-2

Stevenson, R. J., Prescott, J., and Boakes, R. A. (1999). Confusing Tastes and Smells: How Odours Can Influence the Perception of Sweet and Sour Tastes. *Chemical Senses, 24*(6), 627–35. doi:10.1093/chemse/24.6.627

Vazire, S. 2018. Implications of the Credibility Revolution for Productivity, Creativity, and Progress. *Perspectives on Psychological Science, 13*(4), 411–17. doi:10.1177/1745691617751884

Veldhuizen, M. G., Albrecht, J., Zelano, C., Boesveldt, S., Breslin, P., and Lundström, J. N. (2011). Identification of Human Gustatory Cortex by Activation Likelihood Estimation. *Human Brain Mapping*, *32*(12), 2256–66. doi:10.1002/hbm.21188

Ventura, A. K., and Worobey, J. (2013). Early Influences on the Development of Food Preferences. *Current Biology*, *23*(9), R401-408. doi:10.1016/j.cub.2013.02.037

Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L., and Kievit, R. A. (2012). An Agenda for Purely Confirmatory Research. *Perspectives on Psychological Science, 7*(6), 632–38. doi:10.1177/1745691612463078

Winter, B. (2013). Linear Models and Linear Mixed Effects Models in R with Linguistic Applications. *arXiv preprint:1308.5499*

Witt, M., Reutter, K., and Miller, Jr, I. J. (2003). *Morphology of the Peripheral Taste System*. in Handbook of Olfaction and Gustation. Boca Raton, FL, US: CRC Press

World Medical Association. (2013). Ethical principles for medical research involving human subjects. Retrieved from https://www.wma.net/

Wu, A., Dvoryanchikov, G., Pereira, E., Chaudhari, N., and Roper, S. D. (2015). Breadth of Tuning in Taste Afferent Neurons Varies with Stimulus Strength. *Nature Communications, 6*(1):8171. doi:10.1038/ncomms9171

Yarkoni, T. (2020). The Generalizability Crisis. *Behavioral and Brain Sciences*, 1–37. doi:10.1017/S0140525X20001685

Yeomans, M. R., Prescott, J., and Gould, N. J. (2009). Acquired Hedonic and Sensory Characteristics of Odours: Influence of Sweet Liker and Propylthiouracil Taster Status. *The Quarterly Journal of Experimental Psychology, 62*(8), 1648–64. doi:10.1080/17470210802557793